MECHANISTIC STUDIES OF LANTHIPEPTIDE BIOSYNTHESIS

BY

WEIXIN TANG

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Chemistry
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2015

Urbana, Illinois

Doctoral Committee:

Professor Wilfred A. van der Donk, Chair
Professor Satish K. Nair
Professor Chad M. Rienstra
Professor Huimin Zhao

ABSTRACT

Natural products and natural product derivatives have been the leading source of pharmaceutical compounds since the initial application of modern medicine. To fight the increasing occurrence of drug resistance and to reach the ultimate goal of personalized drugs for treating complicated symptoms, novel natural products are demanded. Lanthipeptides are ribosomal peptides with post-translationally incorporated thioether crosslinks named lanthionine. This family of natural products has garnered substantial attention during the past few decades due to their favorable biological activities and the potential for engineering.

This thesis focuses on a subclass of lanthipeptides, the class II compounds, for which the synthesis of dehydroamino acids and the formation of thioether linkages are carried out by a bifunctional lanthionine synthetase. Chapter 2 presents the structural characterization of the enterococcal cytolysin, a lanthipeptide tightly linked to *Enterococcus faecalis* virulence. The stereoselectivity of lanthionine synthesis is discussed in chapters 3 and 4. A non-canonical configuration of lanthionines was discovered in a few lanthipeptides and the selective synthesis of the unusual stereochemistry was found to be induced by the peptide sequence rather than the lanthionine synthetase. Such a substrate-controlled stereoselectivity is rarely identified in naturally occurring enzymatic processes. Although it has been more than ten years since the first *in vitro* reconstitution of a class II lanthionine synthetase, no structural information was available for this class of proteins before the study presented in chapter 5 with respect to the cytolysin synthetase CylM. Unexpectedly, the CylM dehydratase domain resembles the catalytic core of lipid kinases despite the absence of notable sequence homology. Mutagenesis study of CylM provides further insights into the mechanism of the modification process. The maturation of lanthipeptides typically requires a proteolytic step that removes the leader peptide from the modified precursor peptides. Characterization of two peptidases involved in the synthesis of lichenicidin and cytolysin, described in chapters 6 and 7, provides mechanistic insights into these subtilisin-like proteins and reveals their potential as sequence-specific proteases. In addition, structural elucidation of four prochlorosins, a set of lanthipeptides synthesized by a highly substrate-tolerant synthetase ProcM in marine bacteria, is included as chapter 8. The unveiled structural information of these lanthipeptides and the biochemical studies with respect to the biosynthetic process described in this thesis may assist the genome mining and synthetic biology efforts towards novel lanthipeptides for therapeutic purposes and other applications.

For my family

ACKNOWLEDGEMENTS

The journey towards my doctorate has been challenging yet joyful. Besides the scientific discoveries I made, the best things that I have learned over the past five and half years are the courage to face a difficult problem, the ability to search for a solution, and the confidence of myself to find a solution. I have no doubt that I will benefit from all the above experiences throughout my life, both professionally and personally. I want to thank everyone who has been involved in my journey here at University of Illinois at Urbana-Champaign (UIUC). My appreciation is beyond words.

First of all, I want to express my deepest gratitude to my parents, Yuluan Mao and Yichun Tang, who are enjoying their retirement in a small town southeast of China. I would never be who I am without them. They have loved me unconditionally for the past twenty-seven years and there are no words that can possibly express my appreciation. I want to give my special thanks to my boyfriend, Yu Xiang, for his companionship and patience. I know that it is not always easy. Also to my dearest friends since childhood, Ying Zhuang and Feng Gao, for always being there. Fourteen years have passed since we were apart, but our friendship never fades. You are family to me.

I feel extremely lucky for my enrollment at UIUC and for being a member in the research group of Prof. Wilfred van der Donk. As the mentor who guided my initial way to science, Wilfred has provided me everything I could possibly want from a PhD advisor- thoughtful insights about research, freedom to test my ideas, patience and support. I learned how to be a good scientist from him. I want to thank Prof. Huimin Zhao, for serving on my committee and for the helpful and sincere suggestions he has provided all these years. Special thanks also go to Prof. Chad Rienstra for his instructions and advice for my NMR experiments. I appreciate the invitation to his private NMR course, which I still benefit a lot from. I would like to thank Prof. Satish Nair for his involvement in two fruitful collaboration projects with me and for being on my committee. Many thanks go to two previous committee members, Prof. Anne Baranger and Prof. Martin Burke for their help during the early years of my PhD study.

I would like to express my gratitude to Drs. Shi-Hui Dong, Liujie Huo, Lindsay Repka and Neha Garg at UIUC, and Dr. Gonzalo Jiménez-Osés and Prof. K. N. Houk at University of California, Los Angeles (UCLA) for being the best collaborators. I am also very lucky to be

surrounded by fantastic facilities on campus. Special thanks to Furong Sun, Dr. Haijun Yao and Dr. Alexander Ulanov for their help with MS experiments; Drs. Lingyang Zhu, Feng Lin and Jennifer Rapp for their help with NMR experiments.

It is always a pleasure to be a member in the van der land. I would like to thank everyone for making the lab such a nice and warm place, specifically to Martha Freeland and Nan Holda for keeping the lab organized. Special thanks go to Yi Yu, Xiao Yang, Xiling Zhao, Zhengan Zhang, Drs. Min Zeng, Zedu Huang, Jiangtao Gao, Qi Zhang, Huan Wang, Yanxiang Shi and many others for being friends both in and outside the lab. Many thanks to Manny Ortega, Subha Mukherjee, Drs. Mark Walker, Pat Knerr, Gabby Thibodeaux and Noah Bindman for all their helpful discussions. I also want to acknowledge Prof. Bo Li at University of North Carolina at Chapel Hill (UNC) for scientific discussion and career advice, and Prof. Jeff Chan at UIUC for sharing his academic experience with me.

I have spent everyday life with a lot of friends during my stay in Champaign: Yang Wang, Guodong Rao, Pei Wang, Tong Si, Yunzi Luo, Ke Ke, Hang Xing and many others. They make this town a special and memorable place for me. Ultimately, none of my achievements would have been possible without the support from the Chemistry Department of UIUC, and the graduate student service office and the organic and chemical biology area office. Finally, I want to acknowledge the Howard Hughes Medical Institute, the National Institutes of Health and the Ullyott fellowship for supporting my research and education.

TABLE OF CONTENTS

Chapter 1. Lanthipeptides and Their Biosynthesis

1.1 Ribosomally synthesized and post-translationally modified peptides

Natural products and natural product derivatives have been key sources of pharmaceutical compounds over the last century (*1*). Although libraries of large size can be prepared for selection purposes, synthetic compound libraries have been less efficient in identifying lead compounds presumably due to lower structural complexity (*2*). Natural products have been categorized into several subclasses that differ by their biosynthetic routes, including terpenoids, alkaloids, polyketides, and non-ribosomal peptides. As a rapidly expanding family of compounds, ribosomally synthesized and post-translationally modified peptides (RiPPs) have also been recognized as a major family of natural products (**Figure 1.1**) (*3*). The burst of available genomic information has lead to the identification of numerous biosynthetic pathways related to RiPP production, especially in the past decade (*4*). Furthermore, these molecules are produced in all three kingdoms of life and their structures are extremely diverse as a consequence of many different post-translational modifications (**Figure 1.1**) (*5*). Different biological activities have also been assigned to this family of natural products (*6-9*).

Compared to other natural products, RiPPs have several unique and potential favorable features. First, RiPPs are highly evolvable as their precursor peptides are encoded in the genome and synthesized by the ribosome. The biosynthetic pathways responsible for RiPP production are generally short, thus can be more easily identified and manipulated than those of non-ribosomally synthesized compounds (*10*). As vast sequence information becomes available, genome mining for new RiPPs could be extremely fruitful. In addition, derivatives of RiPPs can be accessed by mutating the genes that encode the precursor peptides. These mutated precursor peptides are often still accepted by the corresponding post-translational modification machinery as most of the biosynthetic enzymes have been confirmed to be promiscuous and can tolerate non-natural substrates (*11-14*). As a result, RiPPs have been the focus of scientific research for the past several decades not only for developing new compounds with pharmaceutical potential but also for the novel mechanistic insights hiding behind these post-translational modification processes.

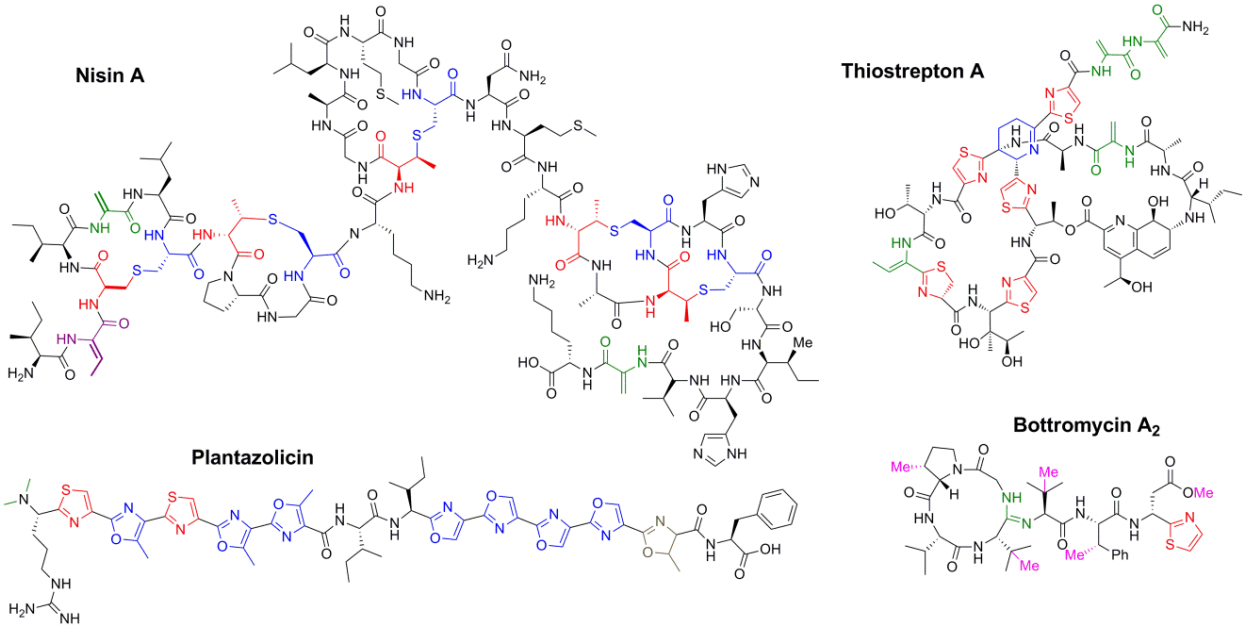

Figure 1.1 Representative members of ribosomally synthesized and post-translationally modified peptides (RiPPs) (*3*). Nisin is one of the longest known RiPPs, which belongs to the lanthipeptide family (*vide infra*) with characteristic lanthionine and methyllanthionine residues (*15*). Plantazolicin is a 14-mer peptide extensively modified by 10 thiazole and (methyl)oxazole heterocycles (*16, 17*). Thiostrepton A is a thiopeptide with post-translationally incorporated pyridinyl macrocycles in addition to thiazole heterocycles and dehydroamino acids (*8*). Bottromycins are a series of related compounds unique due to the presence of a macrocyclic amidine and a decarboxylated thiazole at the C-terminus (*18-21*).

1.2 Lanthipeptides

Lanthipeptides (lanthionine-containing peptides) are one of the best studied classes of the RiPP family (*22*). They are polycyclic peptides characterized by the presence of intramolecular thioether crosslinks. The general biosynthetic route to lanthipeptides is similar to that of most other RiPPs (*22*). A linear precursor peptide is first generated by the ribosome with a leader peptide at the N-terminus that is important for recognition by the post-translational modification machinery. The modifications take place in the core peptide that is located at the C-terminus of the precursor peptide (**Figure 1.2a**) (*22*, *23*). Typically, dehydroamino acids, dehydroalanine (Dha) and dehydrobutyrine (Dhb), synthesized from serine and threonine residues by a dehydratase, respectively (**Figure 1.2a** and **1.2b**), are attacked by cysteine thiols through a Michael-type addition catalyzed by a cyclase. The resulting thioether cross-linked amino acids are termed lanthionine (Lan) and methyllanthionine (MeLan), which define the lanthipeptides (**Figure 1.2a** and **1.2b**). Lanthionine synthetases typically catalyze the formation of (Me)Lan

residues in a highly stereoselective manner, resulting in the formation of a DL configuration ((2*S*, 6*R*)-Lan and (2*S*, 3*S*, 6*R*)-MeLan). The discovery of LL-(Me)Lan ((2*R*, 6*R*) for Lan and (2*R*, 3*R*, 6*R*) for MeLan) in the enterococcal cytolysin serves as the first example of a different stereochemistry for these intramolecular thioether cross-linkages in the lanthipeptide family, which will be discussed in chapter 2. Further characterization revealed that the stereoselectivity is encoded in the substrate sequences rather than dominated by the enzymes (for details, see chapters 3 and 4). Other modifications, such as D-alanine, 3-hydroxy-L-aspartic acid and *S*-aminovinyl D-methylcysteine, may also be incorporated by other enzymes encoded in the lanthipeptide biosynthetic gene cluster (*24*). After installation of the post-translational modifications, the leader peptide is removed by one or more proteases and the mature lanthipeptide is exported outside the producer cells (**Figure 1.2a**) (*25*). Lanthipeptides with antimicrobial activities are termed lantibiotics and some of these exhibit potent growth inhibition activities against various pathogens (*26*, *27*).

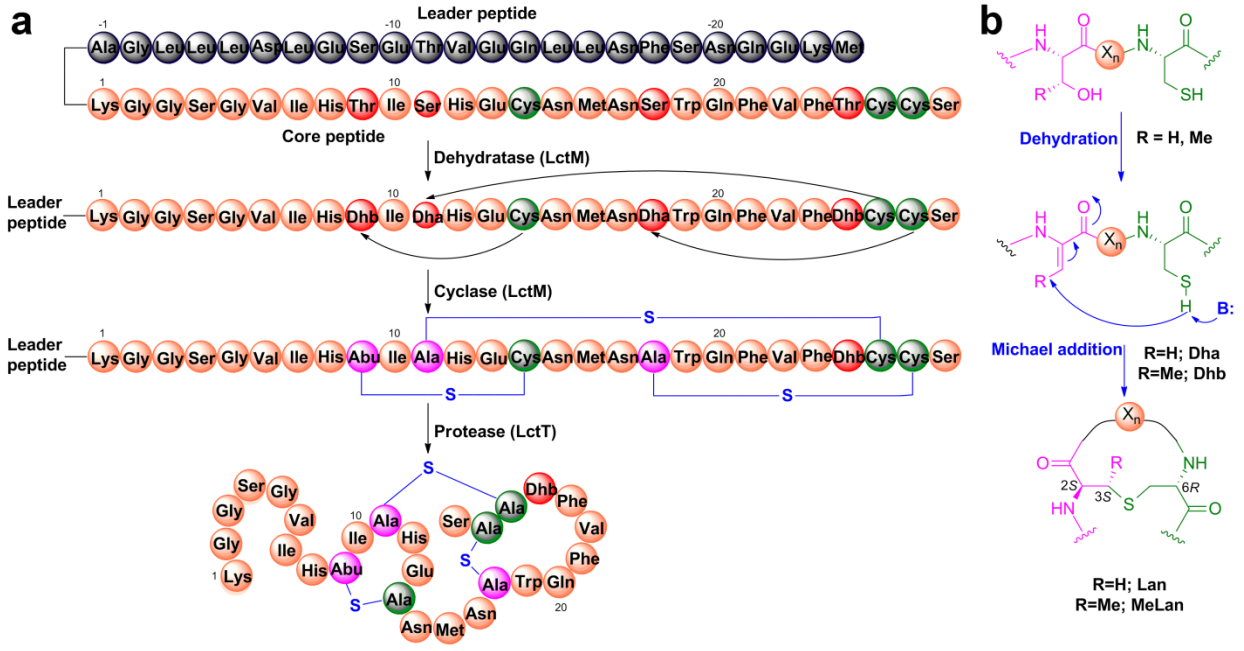

Figure 1.2 Biosynthesis of lanthipeptides. (a) Generic pathway of lanthipeptide biosynthesis using lacticin 481 as an example. The precursor peptide is synthesized by the ribosome with an N-terminal leader peptide and a C-terminal core peptide. The dehydration of Ser and Thr residues to Dha and Dhb residues and the Michael-type addition by Cys residues onto these α,β-unsaturated amino acids are catalyzed by the dehydratase and the cyclase encoded in the lanthipeptide gene cluster, respectively. The leader peptide is then removed by one or more proteases, affording the mature lanthipeptide. **(b)** Dehydration and cyclization reactions that take place during lanthipeptide post-translational modification.

1.3 Four classes of lanthionine synthetases

Lanthipeptides are currently classified into four classes, differentiated by the biosynthetic enzymes responsible for installation of the thioether rings (**Figure 1.3a**) (*22*). For class I lanthipeptides, the dehydration and cyclization reactions are catalyzed by separate enzymes named LanB and LanC (*28-30*), whereas for class II lanthipeptides, the two reactions are carried out by a single bifunctional enzyme LanM (*31*). Class III and IV lanthionine synthetases are trifunctional enzymes with Ser/Thr kinase, phosphoSer/phosphoThr lyase, and cyclase domains (*32, 33*).

Class I dehydratases LanBs are tRNA-dependent enzymes that activate the hydroxyl side chain of Ser/Thr residues by forming an ester bond to facilitate the elimination step (**Figure 1.3b**) (*29*). These proteins are usually about 1,000 amino acids in size and contain two separate domains responsible for ester formation and subsequent elimination, respectively. In particular, NisB, the dehydratase involved in the biosynthesis of the class I lantibiotic nisin, catalyzes the dehydration reaction through Ser/Thr glutamylation and glutamate elimination steps (**Figure 1.3b**) (*28*). The biochemical process has been reconstituted *in vitro* and the catalytic mechanism has been partially revealed by crystallography of the NisB protein (*28, 29*). The cyclization reactions of class I lanthipeptides are carried out by LanC cyclases, which contain a conserved cysteine-cysteine-histidine triad coordinating a zinc ion that is essential for enzyme activity (**Figure 1.3a**) (*30*). LanB and LanC proteins have been reported to interact with each other in cells and catalyze the dehydration and cyclization reactions of their LanA substrates interactively (*34, 35*). The majority of currently known class I lanthipeptides exhibit antimicrobial activities against Gram-positive bacteria, some of which are actively investigated for their potential pharmaceutical utilization (*24*).

The dehydration and cyclization reactions of class II lanthipeptides are carried out by bifunctional synthetases LanMs (**Figure 1.3a**) (*31*). These proteins are typically also about 1,000 amino acids in length with two distinct domains. The N-terminal domain has been assigned to its dehydration activity (*36*). However, the dehydratase domain of LanM proteins shares no homology to the class I LanB dehydratases or any other known proteins in the protein database. Mutagenesis studies of LanM proteins have suggested that they catalyze dehydration reactions through a phosphorylation-elimination mechanism, different from what has been observed for

class I dehydratases (**Figure 1.3b**) (*36*). The C-terminus of LanM proteins is homologous to the class I cyclase NisC, including the conserved zinc-binding ligands that are critical for NisC catalysis. Interestingly, analogs of the dehydratase domain of LanM proteins have been identified in a few biosynthetic pathways responsible for the production of non-lanthipeptide RiPPs, such as polytheonamides (*37*). These LanM dehydratase-like proteins account for the incorporation of dehydro amino acids into the corresponding RiPP products. Different from LanM proteins, these enzymes do not contain a C-terminal cyclase domain, suggesting that the genes corresponding to the dehydratase domain of LanM proteins might have been horizontally transferred and adapted by different biosynthetic machinery to increase the diversity of natural products produced by the producer organisms. No structural information is available for LanM proteins before the work included in this thesis. Characterization of LanMs, focusing on CylM involved in the enterococcal cytolysin biosynthesis, is one of the major projects for my PhD study. The *in vitro* reconstituted activity and crystallization of CylM will be described in chapter 5. CylM is the first and thus far the only LanM with structural information available.

Class III synthetases are trifunctional enzymes with a C-terminal putative cyclase domain that lacks the zinc ligands that are conserved in other lanthipeptide cyclases (**Figure 1.3a**) (*32*). These proteins contain a phosphoserine or phosphothreonine lyase domain at the N-terminus and a Ser/Thr kinase domain located in the middle (**Figure 1.3a**). As a result, class III synthetases are named LanKC (for lanthionine kinase cyclase). Uniquely, LanKC proteins can catalyze the formation of labionin, for which the enolate resulting from the initial Michael-type addition onto the first Dha is not protonated, but instead attacks a second dehydro amino acid to form a carbon-carbon bond. It remains unclear how these cyclizations are carried out as LanKC proteins do not have the conserved zinc ligands that are essential for the activity of the other three classes of synthetases. Class IV synthetases (LanLs) are also trifunctional enzymes that contain an N-terminal phosphoSer/phosphoThr lyase domain and a central kinase domain like class III synthetases (**Figure 1.3a**) (*33*). However, the C-terminal cyclase domain bears homology to the class I cyclase NisC and possesses the zinc ligands (**Figure 1.3a**).

In the four classes of lanthipeptides, three distinct mechanisms have been revealed for the dehydration reaction, whereas at least two catalytic routes have evolved for the cyclization reaction, differentiated by whether the process is zinc-dependent or not. These observations suggest that the formation of dehydro amino acids and (methyl)lanthionine residues may provide

selection privileges to producer organisms, and thus has been independently accessed several times during evolution.

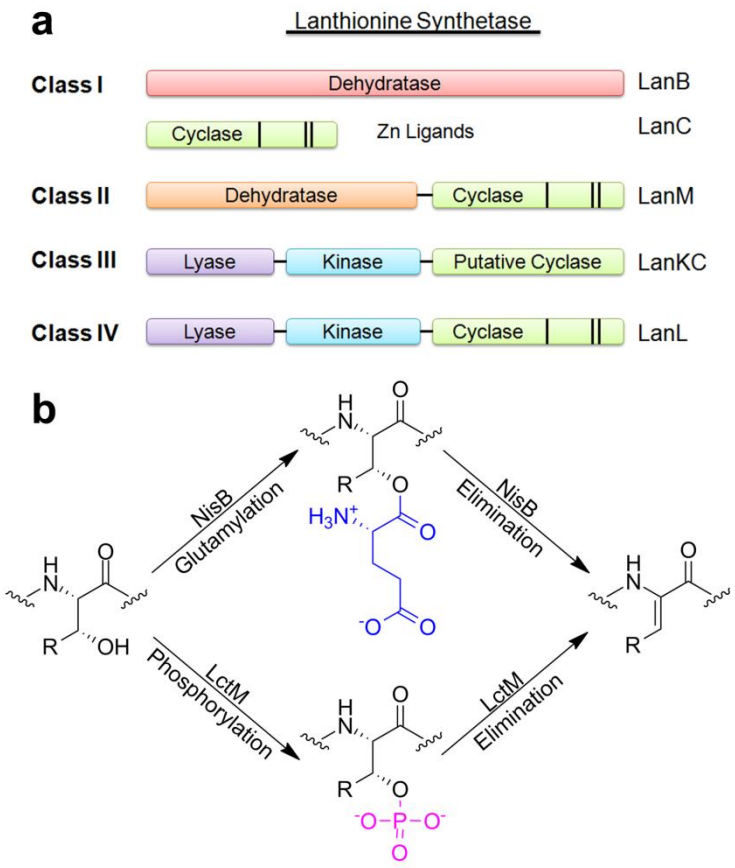

Figure 1.3 Four classes of lanthipeptide synthetases. (a) Schematic representation of four classes of lanthionine synthetases. Zinc ligands are shown as black horizontal bars. **(b)** Different mechanisms of activation adopted by class I and class II dehydratases.

1.4 Proteolytic maturation of lanthipeptides

After the installation of the post-translational modifications, the leader peptides are removed by proteases and the mature lanthipeptides are exported; the leader-containing peptides typically do not exhibit the biological activities of the final mature products (*38*). These proteases are generally encoded by *lanP* or *lanT* genes in the lanthipeptide gene cluster. However, compared to the well-characterized lanthionine synthetases, proteases required for lanthipeptide maturation are much less studied, presumably because these proteins are hard to access with heterologous expression systems.

The maturation of class I lanthipeptides generally requires LanP proteins for leader peptide removal (**Figure 1.4a**) (*24*). These LanP proteins are homologous to the serine protease subtilisin including a conserved catalytic triad (Asp, His and Ser). They have different sizes and cellular locations depending on the presence or absence of an N-terminal secretion signal peptide and a C-terminal cell wall anchor sequence. Knock-out experiments have provided evidence that these LanP proteins are essential for the production of class I lanthipeptides (*40*). However, thus far only three class I LanPs have been heterologously expressed, purified and characterized *in vitro*, including NisP encoded in the nisin gene cluster, EpiP involved in epidermin biosynthesis, and ElxP required for epilancin 15X production (*39, 41-44*). Crystal structures of NisP and a homolog of EpiP have reinforced that these proteins resemble the characteristic conformation and active site geometry of subtilisin-like proteases (*43, 45*). *In vitro* study of ElxP revealed that the sequence spanning the region –5 to +1 of ElxA is important for enzyme recognition, shining light on the substrate specificity of LanP proteins (*39*).

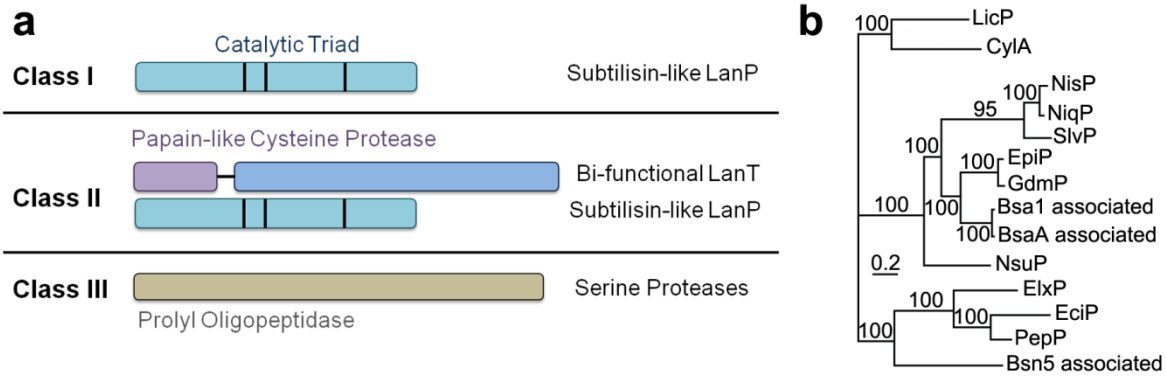

Figure 1.4 Proteases responsible for leader peptide removal of lanthipeptides. (a) Schematic representation of three currently identified types of proteases involved in class I, II and III lanthipeptide biosynthesis. **(b)** Markov chain Monte Carlo (MCMC) phylogenetic tree of LanP enzymes responsible for class I and II lanthipeptide maturation. Two representative class II LanP proteins, LicP and CylA, are included, whereas all others are members of the class I LanP family. Panel B is adapted from a recent published paper focusing on the characterization of the class I LanP ElxP (*39*).

In contrast, leader peptide removal of class II lanthipeptides usually requires a bifunctional LanT transporter protein, which contains an N-terminal papain-like cysteine protease domain that typically cleaves after a double Gly-type motif (GG/GA/GS) (**Figure 1.4a**) (*46, 47*). These LanT proteins maturate lanthipeptides by cleaving off their leader peptides and transporting the peptides into the extracellular medium in an ATP-dependent manner (*48*). As a result, class II

LanT proteins are slightly larger in size than class I transporter proteins due to the presence of the protease domain. Although the leader peptide removal step of most class II lanthipeptides is carried out solely by LanT proteins, a few require one more cleavage process performed by subtilisin-like LanP proteases (**Figure 1.4a**) (*49-52*). These serine proteases are homologous to class I lanthipeptide leader peptidases but are located in a separate clade in a Markov chain Monte Carlo (MCMC) phylogenetic tree (**Figure 1.4b**) (*39*). They appear to remove a short oligopeptide after a LanT protein removes the majority of the leader peptide at a double Gly-type cleavage site. Several examples of LanPs have been identified in class II lanthipeptide biosynthetic pathways, yet few of them have been characterized with respect to either proteolytic activity or substrate specificity, and thus far, no structural information is available for class II LanP proteins. Chapters 6 and 7 will be focusing on biochemical and structural characterization of this subclass of lanthipeptidases.

The proteolytic maturation of class III and class IV lanthipeptides remains largely uncharacterized. Recently, a prolyloligopeptidase-type protease has been linked to the synthesis of the class III lanthipeptide flavipeptin (**Figure 1.4b**) (*53*), suggesting that class III and class IV lanthipeptides may employ different types of proteases for their leader peptide removal.

1.5 Summary and outlook

The research on lanthipeptides has been advanced significantly in the past two decades, resulting in discovery of compounds with favorable biological activities, new post-translational modifications, new classes of biosynthetic machinery, and novel mechanistic insights into enzymes involved in these processes (*22*). The convergent evolution of lanthionine synthetases discussed in section 1.3 indicates that the presence of (methyl)lanthionine residues provides selective advantages for producer organisms to adapt to the environment, as they can produce natural products of high structural diversity at low genetic cost with such post-translational machinery. One particular example, prochlorosins produced by cyanobacteria, will be discussed in chapter 8. Although some lanthipeptides and their biosynthetic routes have been investigated, many more await exploration. For example, post-translational modifications other than dehydration and cyclization can significantly enhance the chemical diversity of lanthipeptides, yet they are underexplored so far. Moreover, only a few class III and class IV lanthipeptides have been identified at present and how LanKC and LanL enzymes install the thioether crosslinks and

labionin residues is currently obscure. Mining novel class III and class IV lanthipeptides as well as mechanistic characterization of how they are synthesized can further expand our knowledge and understanding of lanthipeptides. More importantly, despite the fact that some structural information of lanthionine synthetases is available (*29*, *30*), how the enzymes bind to the substrate remains largely unknown. Such information could be valuable for the engineering efforts of lanthipeptides (*54*). In an era when massive sequence and genomic information becomes available, I am expecting more and more novel RiPPs and lanthipeptides to be identified using modern genome-mining strategies in the future, continuously advancing our understanding of these biological systems and enriching our collections of natural products.

1.6 References

1. Newman, D. J., and Cragg, G. M. (2007) Natural products as sources of new drugs over the last 25 years, *J. Nat. Pro. 70*, 461-477.
2. Payne, D. J., Gwynn, M. N., Holmes, D. J., and Pompliano, D. L. (2007) Drugs for bad bugs: confronting the challenges of antibacterial discovery, *Nat. Rev. Drug Discov. 6*, 29-40.
3. Arnison, P. G., Bibb, M. J., Bierbaum, G., Bowers, A. A., Bugni, T. S., Bulaj, G., Camarero, J. A., Campopiano, D. J., Challis, G. L., Clardy, J., Cotter, P. D., Craik, D. J., Dawson, M., Dittmann, E., Donadio, S., Dorrestein, P. C., Entian, K.-D., Fischbach, M. A., Garavelli, J. S., Göransson, U., Gruber, C. W., Haft, D. H., Hemscheidt, T. K., Hertweck, C., Hill, C., Horswill, A. R., Jaspars, M., Kelly, W. L., Klinman, J. P., Kuipers, O. P., Link, A. J., Liu, W., Marahiel, M. A., Mitchell, D. A., Moll, G. N., Moore, B. S., Müller, R., Nair, S. K., Nes, I. F., Norris, G. E., Olivera, B. M., Onaka, H., Patchett, M. L., Piel, J., Reaney, M. J. T., Rebuffat, S., Ross, R. P., Sahl, H.-G., Schmidt, E. W., Selsted, M. E., Severinov, K., Shen, B., Sivonen, K., Smith, L., Stein, T., Süssmuth, R. E., Tagg, J. R., Tang, G.-L., Truman, A. W., Vederas, J. C., Walsh, C. T., Walton, J. D., Wenzel, S. C., Willey, J. M., and van der Donk, W. A. (2013) Ribosomally Synthesized and Post-translationally Modified Peptide Natural Products: Overview and Recommendations for a Universal Nomenclature, *Nat. Prod. Rep. 30*, 108-160.
4. Velásquez, J. E., and van der Donk, W. A. (2011) Genome mining for ribosomally synthesized natural products, *Curr. Opin. Chem. Biol. 15*, 11-21.
5. McIntosh, J. A., Donia, M. S., and Schmidt, E. W. (2009) Ribosomal peptide natural products: bridging the ribosomal and nonribosomal worlds, *Nat. Prod. Rep. 26*, 537–559.
6. Klaenhammer, T. R. (1993) Genetics of bacteriocins produced by lactic acid bacteria, *FEMS Microbiol. Rev. 12*, 39-85.
7. Roy, R. S., Gehring, A. M., Milne, J. C., Belshaw, P. J., and Walsh, C. T. (1999) Thiazole and oxazole peptides: biosynthesis and molecular machinery, *Nat. Prod. Rep. 16*, 249-263.
8. Kelly, W. L., Pan, L., and Li, C. (2009) Thiostrepton biosynthesis: prototype for a new family of bacteriocins, *J. Am. Chem. Soc. 131*, 4327-4334.
9. Houssen, W. E., Koehnke, J., Zollman, D., Vendome, J., Raab, A., Smith, M. C., Naismith, J. H., and Jaspars, M. (2012) The discovery of new cyanobactins from cyanothece PCC 7425 defines a new signature for processing of patellamides, *ChemBioChem 13*, 2683-2689.
10. Yang, X., and van der Donk, W. A. (2013) Ribosomally Synthesized and Post-Translationally Modified Peptide Natural Products: New Insights into the Role of Leader and Core Peptides during Biosynthesis, *Chem. Eur. J. 19*, 7662-7677.

11. Levengood, M. R., Knerr, P. J., Oman, T. J., and van der Donk, W. A. (2009) In vitro mutasynthesis of lantibiotic analogues containing nonproteinogenic amino acids, *J. Am. Chem. Soc. 131*, 12024-12025.
12. Mitchell, D. A., Lee, S. W., Pence, M. A., Markley, A. L., Limm, J. D., Nizet, V., and Dixon, J. E. (2009) Structural and functional dissection of the heterocyclic peptide cytotoxin streptolysin S, *J. Biol. Chem. 284*, 13004-13012.
13. Bowers, A. A., Acker, M. G., Koglin, A., and Walsh, C. T. (2010) Manipulation of thiocillin variants by prepeptide gene replacement: structure, conformation, and activity of heterocycle substitution mutants, *J. Am. Chem. Soc. 132*, 7519-7527.
14. Knappe, T. A., Manzenrieder, F., Mas-Moruno, C., Linne, U., Sasse, F., Kessler, H., Xie, X., and Marahiel, M. A. (2011) Introducing lasso peptides as molecular scaffolds for drug design: engineering of an integrin antagonist, *Angew. Chem. Int. Ed. 50*, 8714-8717.
15. Gross, E., and Morell, J. L. (1971) The Structure of Nisin, *J. Am. Chem. Soc. 93*, 4634-4635.
16. Kalyon, B., Helaly, S. E., Scholz, R., Nachtigall, J., Vater, J., Borriss, R., and Süssmuth, R. D. (2011) Plantazolicin A and B: structure elucidation of ribosomally synthesized thiazole/oxazole peptides from Bacillus amyloliquefaciens FZB42, *Org. Lett. 13*, 2996-2999.
17. Molohon, K. J., Melby, J. O., Lee, J., Evans, B. S., Dunbar, K. L., Bumpus, S. B., Kelleher, N. L., and Mitchell, D. A. (2011) Structure Determination and Interception of Biosynthetic Intermediates for the Plantazolicin Class of Highly Discriminating Antibiotics, *ACS Chem. Biol. 6*, 1307-1313.
18. Huo, L., Rachid, S., Stadler, M., Wenzel, S. C., and Müller, R. (2012) Synthetic Biotechnology to Study and Engineer Ribosomal Bottromycin Biosynthesis, *Chem. Biol. 19*, 1278-1287.
19. Crone, W. J. K., Leeper, F. J., and Truman, A. W. (2012) Identification and characterisation of the gene cluster for the anti-MRSA antibiotic bottromycin:expanding the biosynthetic diversity of ribosomal peptides, *Chem. Sci. 3*, 3516-3521.
20. Gomez-Escribano, J. P., Song, L., Bibb, M. J., and Challis, G. L. (2012) Posttranslational b-methylation and macrolactamidination in the biosynthesis of the bottromycin complex of ribosomal peptide antibiotics, *Chem. Sci. 3*, 3522-3525.
21. Hou, Y., Tianero, M. D., Kwan, J. C., Wyche, T. P., Michel, C. R., Ellis, G. A., Vazquez-Rivera, E., Braun, D. R., Rose, W. E., Schmidt, E. W., and Bugni, T. S. (2012) Structure and Biosynthesis of the Antibiotic Bottromycin D, *Org. Lett. 14*, 5050-5053.
22. Knerr, P. J., and van der Donk, W. A. (2012) Discovery, biosynthesis, and engineering of lantipeptides, *Annu. Rev. Biochem. 81*, 479-505.
23. Bierbaum, G., and Sahl, H. G. (2009) Lantibiotics: mode of action, biosynthesis and bioengineering, *Curr. Pharm. Biotechnol. 10*, 2-18.

24. Chatterjee, C., Paul, M., Xie, L., and van der Donk, W. A. (2005) Biosynthesis and mode of action of lantibiotics, *Chem. Rev. 105*, 633-684.
25. Plat, A., Kuipers, A., Rink, R., and Moll, G. N. (2013) Mechanistic Aspects of Lanthipeptide Leaders, *Curr. Protein Pept. Sci. 14*, 85.
26. Cotter, P. D., Hill, C., and Ross, R. P. (2005) Bacterial lantibiotics: strategies to improve therapeutic potential, *Curr. Protein Pept. Sci. 6*, 61-75.
27. Nishie, M., Nagao, J. I., and Sonomoto, K. (2012) Antibacterial peptides "bacteriocins": an overview of their diverse characteristics and applications, *Biocontrol Sci. 17*, 1-16.
28. Garg, N., Salazar-Ocampo, L. M., and van der Donk, W. A. (2013) In vitro activity of the nisin dehydratase NisB, *Proc. Natl. Acad. Sci. U. S. A. 110*, 7258-7263.
29. Ortega, M. A., Hao, Y., Zhang, Q., Walker, M. C., van der Donk, W. A., and Nair, S. K. (2015) Structure and mechanism of the tRNA-dependent lantibiotic dehydratase NisB, *Nature 517*, 509-512.
30. Li, B., Yu, J. P., Brunzelle, J. S., Moll, G. N., van der Donk, W. A., and Nair, S. K. (2006) Structure and mechanism of the lantibiotic cyclase involved in nisin biosynthesis, *Science 311*, 1464-1467.
31. Xie, L., Miller, L. M., Chatterjee, C., Averin, O., Kelleher, N. L., and van der Donk, W. A. (2004) Lacticin 481: in vitro reconstitution of lantibiotic synthetase activity, *Science 303*, 679-681.
32. Müller, W. M., Schmiederer, T., Ensle, P., and Süssmuth, R. D. (2010) In vitro biosynthesis of the prepeptide of type-III lantibiotic labyrinthopeptin A2 including formation of a C-C bond as a post-translational modification, *Angew. Chem., Int. Ed. 49*, 2436-2440.
33. Goto, Y., Li, B., Claesen, J., Shi, Y., Bibb, M. J., and van der Donk, W. A. (2010) Discovery of unique lanthionine synthetases reveals new mechanistic and evolutionary insights, *PLoS Biol. 8*, e1000339.
34. Siegers, K., Heinzmann, S., and Entian, K.-D. (1996) Biosynthesis of lantibiotic nisin. Posttranslational modification of its prepeptide occurs at a multimeric membrane-associated lanthionine synthetase complex., *J. Biol. Chem. 271*, 12294-12301.
35. Lubelski, J., Khusainov, R., and Kuipers, O. P. (2009) Directionality and Coordination of Dehydration and Ring Formation during Biosynthesis of the Lantibiotic Nisin, *J. Biol. Chem. 284*, 25962-25972.
36. You, Y. O., and van der Donk, W. A. (2007) Mechanistic investigations of the dehydration reaction of lacticin 481 synthetase using site-directed mutagenesis, *Biochemistry 46*, 5991-6000.
37. Freeman, M. F., Gurgui, C., Helf, M. J., Morinaka, B. I., Uria, A. R., Oldham, N. J., Sahl, H. G., Matsunaga, S., and Piel, J. (2012) Metagenome mining reveals polytheonamides as posttranslationally modified ribosomal peptides, *Science 338*, 387-390.

38. Oman, T. J., and van der Donk, W. A. (2010) Follow the leader: the use of leader peptides to guide natural product biosynthesis, *Nat. Chem. Biol. 6*, 9-18.
39. Ortega, M. A., Velásquez, J. E., Garg, N., Zhang, Q., Joyce, R. E., Nair, S. K., and van der Donk, W. A. (2014) Substrate Specificity of the Lanthipeptide Peptidase ElxP and the Oxidoreductase ElxO, *ACS Chem. Biol. 9*, 1718-1725.
40. van der Meer, J. R., Polman, J., Beerthuyzen, M. M., Siezen, R. J., Kuipers, O. P., and de Vos, W. M. (1993) Characterization of the *Lactococcus lactis* nisin A operon genes *nisP*, encoding a subtilisin-like serine protease involved in precursor processing, and *nisR*, encoding a regulatory protein involved in nisin biosynthesis, *J. Bacteriol. 175*, 2578-2588.
41. Geissler, S., Götz, F., and Kupke, T. (1996) Serine protease EpiP from *Staphylococcus epidermidis* catalyzes the processing of the epidermin precursor peptide, *J. Bacteriol. 178*, 284-288.
42. Abts, A., Montalban-Lopez, M., Kuipers, O. P., Smits, S. H., and Schmitt, L. (2013) NisC binds the FxLx motif of the nisin leader peptide, *Biochemistry 52*, 5387-5395.
43. Xu, Y., Li, X., Li, R., Li, S., Ni, H., Wang, H., Xu, H., Zhou, W., Saris, P. E., Yang, W., Qiao, M., and Rao, Z. (2014) Structure of the nisin leader peptidase NisP revealing a C-terminal autocleavage activity, *Acta Crystallogr. Sect. D Biol. Crystallogr. 70*, 1499-1505.
44. Velásquez, J. E., Zhang, X., and van der Donk, W. A. (2011) Biosynthesis of the Antimicrobial Peptide Epilancin 15X and its Unusual N-terminal Lactate Moiety, *Chem. Biol. 18*, 857-867.
45. Kuhn, M. L., Prachi, P., Minasov, G., Shuvalova, L., Ruan, J., Dubrovska, I., Winsor, J., Giraldi, M., Biagini, M., Liberatori, S., Savino, S., Bagnoli, F., Anderson, W. F., and Grandi, G. (2014) Structure and protective efficacy of the *Staphylococcus aureus* autocleaving protease EpiP, *FASEB J. 28*, 1780-1793.
46. Furgerson Ihnken, L. A., Chatterjee, C., and van der Donk, W. A. (2008) *In vitro* Reconstitution and Substrate Specificity of a Lantibiotic Protease, *Biochemistry 47*, 7352-7363.
47. Nishie, M., Shioya, K., Nagao, J., Jikuya, H., and Sonomoto, K. (2009) ATP-dependent leader peptide cleavage by NukT, a bifunctional ABC transporter, during lantibiotic biosynthesis, *J. Biosci. Bioeng. 108*, 460-464.
48. Nishie, M., Sasaki, M., Nagao, J., Zendo, T., Nakayama, J., and Sonomoto, K. (2011) Lantibiotic transporter requires cooperative functioning of the peptidase domain and the ATP binding domain, *J. Biol. Chem. 286*, 11163-11169.
49. Booth, M. C., Bogie, C. P., Sahl, H.-G., Siezen, R. J., Hatter, K. L., and Gilmore, M. S. (1996) Structural analysis and proteolytic activation of Enterococcus faecalis cytolysin, a novel lantibiotic, *Mol. Microbiol. 21*, 1175-1184.

50. Caetano, T., Krawczyk, J. M., Mosker, E., Süssmuth, R. D., and Mendo, S. (2011) Heterologous expression, biosynthesis, and mutagenesis of type II lantibiotics from *Bacillus licheniformis* in *Escherichia coli*, *Chem. Biol. 18*, 90-100.
51. Wang, J., Zhang, L., Teng, K., Sun, S., Sun, Z., and Zhong, J. (2014) Cerecidins, novel lantibiotics from *Bacillus cereus* with potent antimicrobial activity, *Appl. Environ. Microbiol. 80*, 2633-2643.
52. Lohans, C. T., Li, J. L., and Vederas, J. C. (2014) Structure and biosynthesis of carnolysin, a homologue of enterococcal cytolysin with D-amino acids, *J. Am. Chem. Soc. 136*, 13150-13153.
53. Völler, G. H., Krawczyk, B., Ensle, P., and Süssmuth, R. D. (2013) Involvement and unusual substrate specificity of a prolyl oligopeptidase in class III lanthipeptide maturation, *J. Am. Chem. Soc. 135*, 7426-7429.
54. Oman, T. J., Knerr, P. J., Bindman, N. A., Velasquez, J. E., and van der Donk, W. A. (2012) An engineered lantibiotic synthetase that does not require a leader peptide on its substrate, *J. Am. Chem. Soc. 134*, 6952–6955.

Chapter 2. Structural Characterization of the Enterococcal Cytolysin Reveals Unusual Lanthionine Stereochemistry

2.1 Introduction

The enterococcal cytolysin is a two-component lantibiotic made up of the post-translationally modified peptides $CylL_L$" and $CylL_S$" (cytolysin L and S). It is produced by many clinical isolates of *Enterococcus faecalis* and is toxic to a broad range of gram positive bacteria (*1-3*). The enterococcal cytolysin is a unique lantibiotic in that it has hemolytic activity against eukaryotic cells in addition to its antimicrobial activity (*2, 4*). The production of cytolysin enhances virulence in infection models and its association with acute patient mortality is supported by epidemiological data (*5-7*). Despite the intensive research focused on the biological functions and biosynthesis of the enterococcal cytolysin (*8-10*), its structure has never been reported until the work described in this chapter.

Like other members in the lantibiotic family (*11*), each component of the enterococcal cytolysin is initially expressed as a linear precursor peptide ($CylL_L$ and $CylL_S$) consisting of an N-terminal leader peptide and a C-terminal core peptide. The lanthionine synthetase CylM then dehydrates Ser and Thr residues in the core peptide followed by Michael-type addition of the thiols of Cys residues to the resulting dehydroamino acids to generate lanthionine (Lan) and methyllanthionine (MeLan) structures, respectively (*8, 12*). CylB then removes most of the leader peptide and secretes the cyclized peptides (*13*). Extracellularly, six additional residues are removed from the N-terminus of each peptide by CylA to generate $CylL_L$" and $CylL_S$", which make up mature cytolysin (*8, 14, 15*).

Reproduced in part with the permission from:
Tang, W., and van der Donk, W. A. (2013) The sequence of the enterococcal cytolysin imparts unusual lanthionine stereochemistry, *Nat. Chem. Biol. 9*, 157-159.

Intrigued by the unique properties of the enterococcal cytolysin and its contribution to the virulence of *E. faecalis*, I set out to characterize its structure by determining the ring topology and the stereochemistry of its (Me)Lan residues. Because high level production of cytolysin in *E. faecalis* requires contact with target mammalian cells (*10*), only relatively small amounts of material can be obtained, which probably has contributed to the lack of prior structural studies.

2.2 Results

In this work, $CylL_L$" and $CylL_S$" were both produced in *E. coli*. First, $cylL_L$ and $cylL_S$ encoding the precursor peptides as well as *cylM* encoding the lanthionine synthetase were synthesized with codon optimization for use in *E. coli*. Each precursor peptide was co-expressed as an N-terminal His_6-tag fusion peptide with untagged CylM from a pRSF-Duet-1 vector, as previously reported for a few other lantibiotics (*16*). After purification by immobilized metal affinity chromatography, the CylM-modified $CylL_L$ and $CylL_S$ exhibited mass losses of approximately 130 Da and 72 Da, respectively. The exact mass shift was confirmed by subsequent endoproteinase AspN digest to remove the first 25 and 37 residues of the leader peptides of $CylL_L$ and $CylL_S$, respectively (**Figure 2.1**). Mass spectrometric analysis of the modified $CylL_L$ indicated that the major product had undergone 7 dehydrations, whereas purified CylM-modified $CylL_S$ treated with AspN displayed a mass consistent with 4 dehydrations (**Figure 2.1**). These results match the reported masses of $CylL_L$" and $CylL_S$", for which 7 and 4 dehydrations were suggested, respectively (*15*). To verify the formation of the thioether linkages and resolve the intramolecular ring topology, tandem mass spectrometry (MS) was employed to analyze the fragmentation pattern of the 7-fold dehydrated $CylL_L$ and 4-fold dehydrated $CylL_S$ (**Figure 2.2**). Fragmentation was suppressed at stretches of sequence comprising residues 1-5, 14-18, and 34-38 for the core peptide of modified $CylL_L$, suggesting 3 rings were present in these regions. Similarly, suppressed fragmentation was detected between residues 1-5 and 17-21 of modified $CylL_S$, suggesting it had 2 rings spanning these residues. The fragmentation pattern of modified $CylL_L$ also suggested that Thr27 escaped dehydration by CylM. Collectively, these data suggest the structures shown in **Figure 2.2** for modified $CylL_L$ and $CylL_S$, respectively.

The AspN digestion resulted in smaller fragments that allowed tandem MS analysis but did not afford the mature peptides because five residues of the leader peptide remained. To obtain

mature $CylL_L$" and $CylL_S$", a Glu−1Lys mutation was introduced in $CylL_L$ and $CylL_S$ for subsequent leader peptide removal with trypsin. After co-expression with CylM, His_6-$CylL_L$-E−1K exhibited a mass loss consistent with 7 dehydrations and $CylL_S$-E−1K displayed a mass consistent with loss of 4 water molecules. Tandem MS analysis verified that modified $CylL_L$-E−1K and $CylL_S$-E−1K displayed the same fragmentation patterns as those of the modified wild type peptides (data not shown). $CylL_L$" and $CylL_S$" were successfully obtained by digesting CylM-modified $CylL_L$-E−1K and $CylL_S$-E−1K with trypsin followed by HPLC purification (**Figures 2.3c** and **2.4c**).

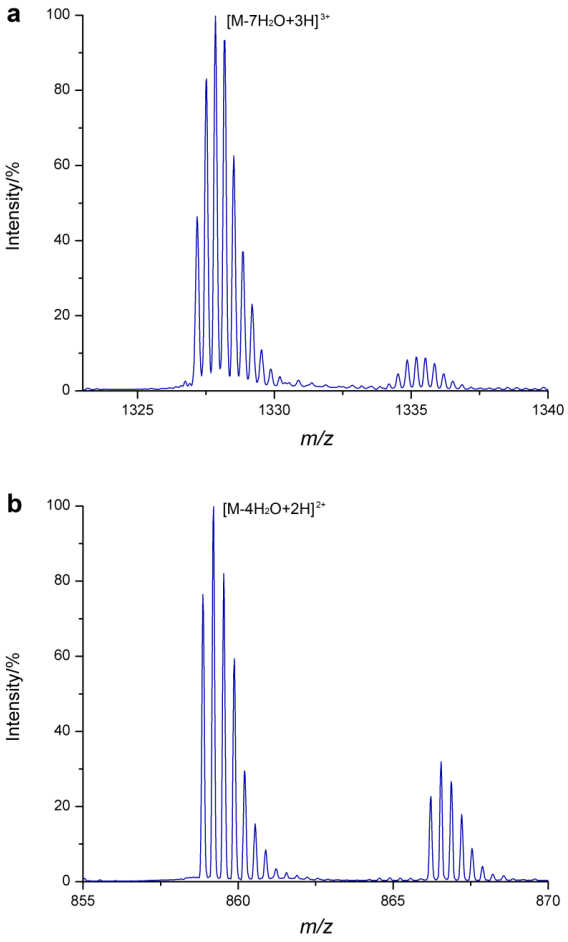

Figure 2.1 ESI-Q/TOF mass spectra of (a) 7-fold dehydrated $CylL_L$, and (b) 4-fold dehydrated $CylL_S$ core peptides modified by CylM in *E. coli* and digested by AspN. $[M+3H]^{3+}$ is shown for the $CylL_L$ core peptides, and $[M+2H]^{2+}$ is shown for the $CylL_S$ core peptide. Both core peptides contain five additional amino acids at their N-terminus originating from the leader peptides as a result of AspN cleavage. The sequences of the leader peptides of $CylL_L$ and $CylL_S$ are MENLSVVPSFEELSVEEMEAIQGSGDVQAE and MLNKENQENYSNKLELVGPSFEELSLEEMEAIQGSGDVQAE, respectively. The leader peptide residues remaining after cleavage with AspN are underlined.

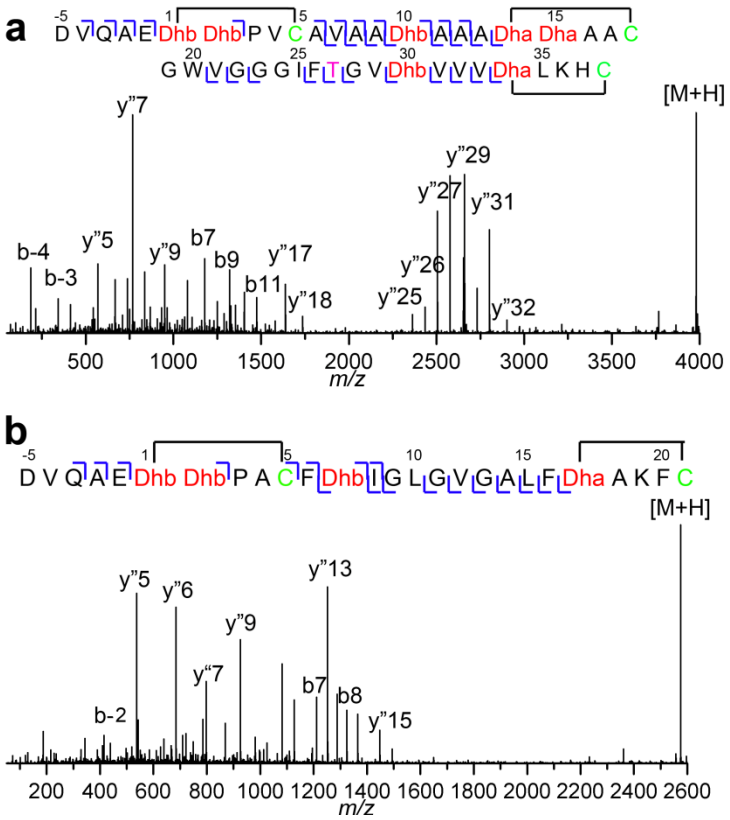

Figure 2.2 Tandem MS analysis of CylM-modified CylL$_L$ (a) and CylL$_S$ (b) produced in *E. coli*. Both core peptides contain five residues at their N-terminus that remain from the leader peptides after AspN cleavage. The b2 fragment ions observed for modified CylL$_L$ and CylL$_S$ are attributed to in-ring fragmentation. Similar fragmentation has been observed in the A rings of nisin and geobacillin II (*16, 17*). Dha, dehydroalanine; Dhb, dehydrobutyrine.

Although the analysis of the tandem MS fragmentation data suggests that the A rings in both CylL$_L$" and CylL$_S$" are formed between Dhb1 and Cys5, it is difficult to rule out by tandem MS alone that Cys5 is not linked to Dhb2. If this were the case, the final products should have an N-terminal 2-oxobutyryl group as a result of spontaneous hydrolysis of Dhb1, and the final product should display a 1 Da mass increase. To confirm the proposed topology of the A ring, CylL$_L$" and CylL$_S$" were analyzed by high resolution electrospray ionization MS (*18*). The observed masses of CylL$_L$" and CylL$_S$" were 3435.6870 Da and 2031.0260 Da, respectively, consistent with structures containing a MeLan at their N-terminus with calculated masses of 3435.7059 and 2031.0122 (errors of 5.5 ppm and 6.8 ppm, respectively, **Figures 2.3** and **2.4**). The data are inconsistent with 2-oxobutyryl as the N-terminal residue. Therefore, the A rings of CylL$_L$" and CylL$_S$" were unambiguously assigned to MeLan residues formed from Cys5 and Dhb1.

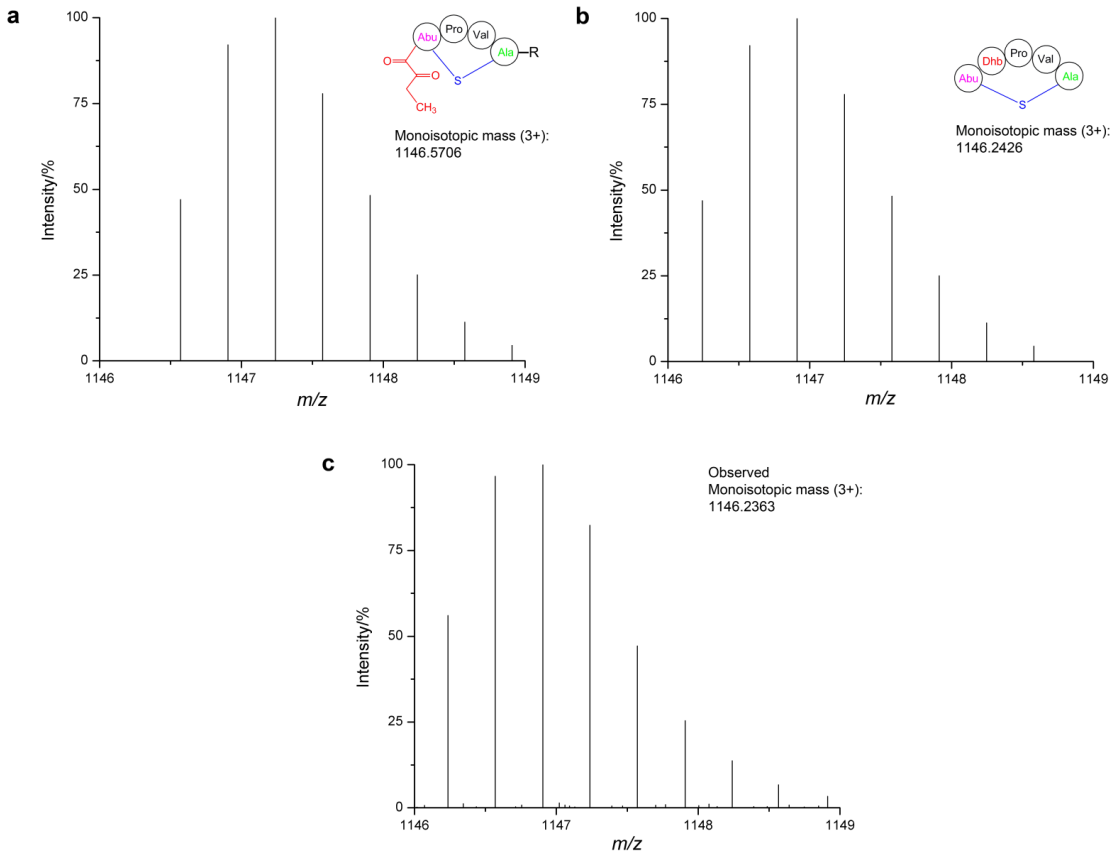

Figure 2.3 Simulated mass spectra of CylL_L" with a 2-oxobutyryl group (a) or with a MeLan group (b) at the N-terminus, and comparison with the observed ESI-Q/TOF mass spectrum of CylL_L" (c). Note that the 3+ ions are shown and hence that the calculated mass difference is correspondingly smaller for the structures drawn (i.e. mass difference is less than 1 Da).

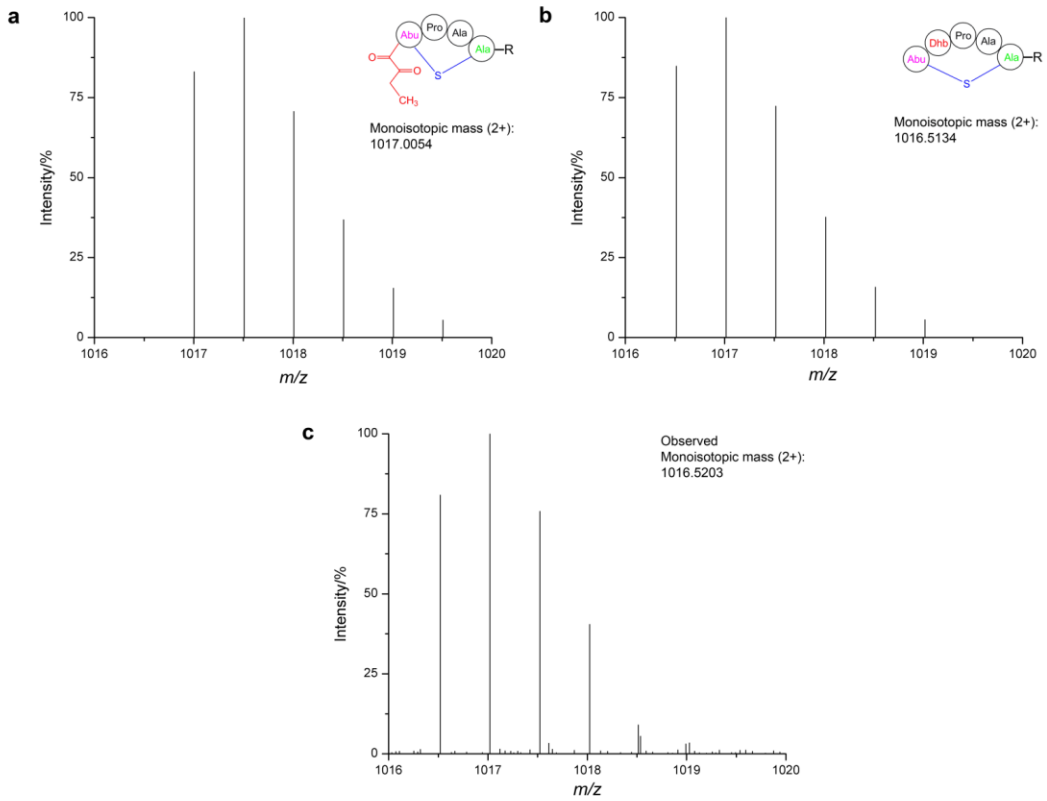

Figure 2.4 Simulated mass spectra of CylL$_S$" with a 2-oxobutyryl group (a) or with a MeLan group (b) at the N-terminus, and comparison with the observed ESI-Q/TOF mass spectrum of CylL$_S$" (c). Note that the 2+ ions are shown and hence that the calculated mass difference is correspondingly smaller for the structures drawn (i.e. expected mass difference is less than 1 Da).

Based on the characterization of select lantibiotics in previous studies (*19*), all the enzymatically formed Lan and MeLan residues in lantibiotics are generally believed to exhibit a "DL" configuration, i.e. the D-configuration at the α-carbon of the former Ser/Thr residue and the L-configuration at the α-carbon of the former Cys residue ((2*S*,6*R*) for Lan and (2*S*,3*S*,6*R*) for MeLan, **Figure 2.6a**, left panel). To confirm the expected stereochemistry of the (Me)Lan residues in CylL$_L$" and CylL$_S$", the peptides were hydrolyzed in 6 M HCl, derivatized to their corresponding pentafluoropropionamide methyl esters, and analyzed by GC-MS with a chiral stationary phase. Lan and MeLan standards of different stereochemistry were synthesized and derivatized using previously reported methods (*20*, *21*). A Chirasil-L-Val coated GC column was used for analysis and the GC trace was monitored for a characteristic fragment of 365 Da for Lan and 379 Da for MeLan residues using electron-impact mass spectrometry. **Figure 2.5** displays the GC-MS traces for the three Lan and two MeLan standards after deivatization. Unexpectedly, the

MeLan residues in both CylM-modified $CylL_L$ and $CylL_S$ had the (2*R*,3*R*,6*R*) configuration (**Figure 2.6a**, right panel; **Figures 2.6b** and **2.6c**, left panels) by comparison with synthetic standards (i.e. both alpha carbons have the L-configuration, hereafter referred to as LL). These findings indicate that the Michael-type addition still occurs with overall *anti* stereochemistry, but with opposite face selectivity for both the attack of the Cys nucleophile onto the Dhb and the protonation of the resulting enolate. This observation constitutes the first example of alternative stereochemistry in lantibiotics. Perhaps even more surprising is the finding that the Lan residue in modified $CylL_S$ exhibits the traditional *meso* configuration (DL) (**Figure 2.6a**, left panel), and that the two Lan residues in modified $CylL_L$ are present in a 1:1 ratio of DL and LL configurations (**Figures 2.6b** and **2.6c**, right panels). These data show that CylM can catalyze the Michael-type addition with different stereochemistries in a single polypeptide substrate. The 1:1 ratio of DL and LL configurations of the Lan in modified $CylL_L$ could be either a mixture of stereoisomers for both the B and C rings, or it could be a consequence of one ring having the DL configuration and the other ring having the LL configuration. To distinguish between these possibilities, the mutants $CylL_L$-S14T and $CylL_L$-S15T were co-expressed with CylM in *E. coli*, purified, hydrolyzed, derivatized, and analyzed by chiral GC-MS (**Figures 2.9** and **2.10**). The modified $CylL_L$-S14T was found to contain LL-MeLan and only DL-Lan whereas CylM-modified $CylL_L$-S15T exhibited a similar GC trace as the modified wild type $CylL_L$ (LL-MeLan, and a 1:1 ratio of LL-Lan and DL-Lan). These results suggest that the B ring of $CylL_L$ is formed between Ser14 and Cys18 and has a LL configuration, whereas the C ring has the DL configuration. With this information, the final structures of $CylL_L$'' and $CylL_S$'' are established (**Figure 2.11**).

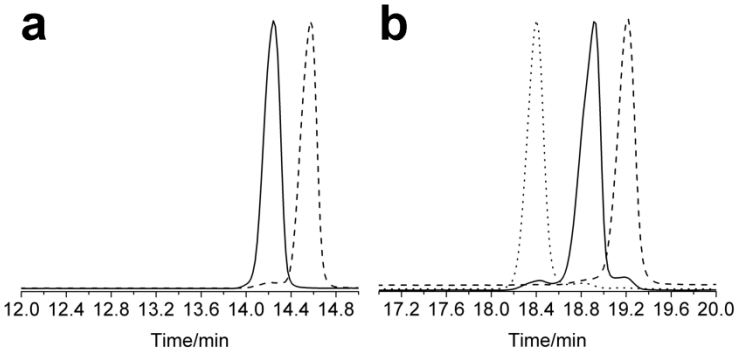

Figure 2.5 GC-MS traces (selected ion monitoring, SIM, at 365 Da for Lan and 379 Da for MeLan) of derivatized (Me)Lan standards. **(a)** Derivatized DL-MeLan (solid line) and LL-MeLan standards (dashed line). **(b)** Derivatized DD-Lan (short dashed line), DL-Lan (solid line) and LL-Lan (long dashed line) standards.

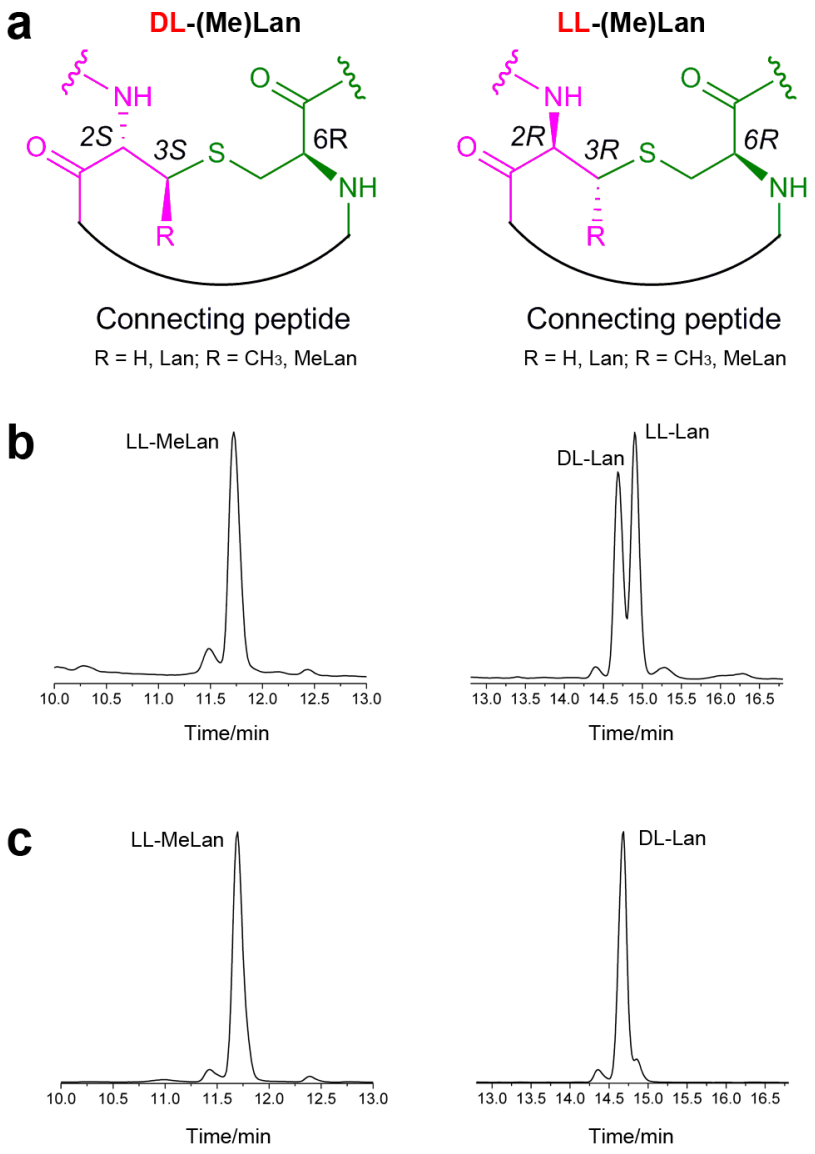

Figure 2.6 GC-MS analysis of CylM-modified CylL$_L$ and CylL$_S$. (a) Canonical (DL, left) and unusual (LL, right) stereochemistries of (Me)Lan residues in previously characterized lantibiotics and in the enterococcal cytolysin, respectively. **(b)** GC-MS traces of hydrolyzed and derivatized MeLan and Lan residues from CylM-modified CylL$_L$. **(c)** GC-MS traces of hydrolyzed and derivatized MeLan and Lan residues from CylM-modified CylL$_S$. Selected ion chromatogram is shown by extracting the total ion chromatogram with characteristic fragment masses of 379 Da for MeLan and 365 Da for Lan, respectively. Stereochemistries of MeLan and Lan residues were assigned by co-injection with synthetic standards derivatized using the same procedure (**Figures 2.7** and **2.8**). For preparation of the (Me)Lan standards, see section 2.3.

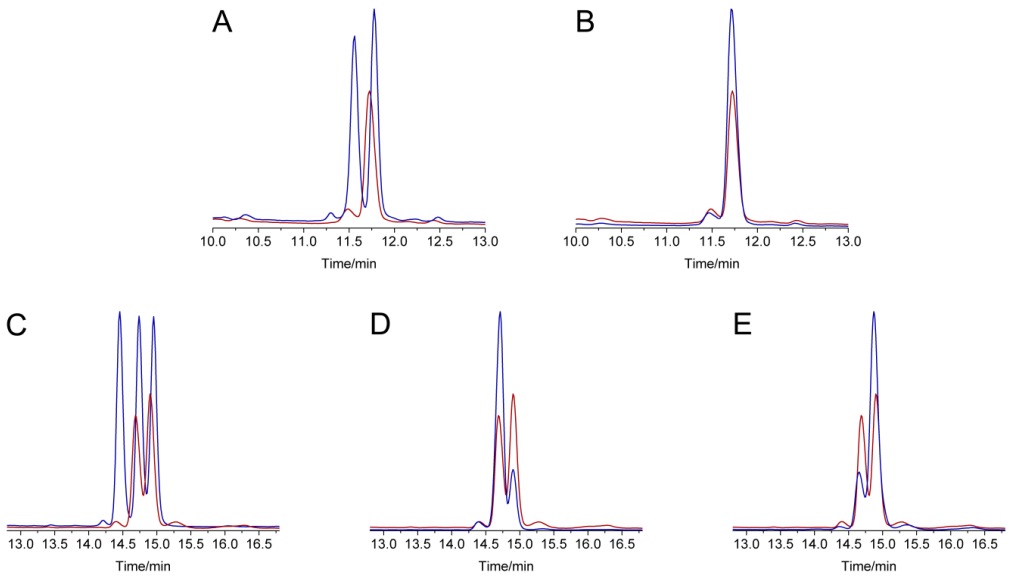

Figure 2.7 GC-MS traces for co-injections of synthetic, derivatized Lan/MeLan standards and hydrolyzed/derivatized Lan/MeLan residues obtained from *in vivo* modified CylL$_L$ (selected ion monitoring, SIM, at 365 Da for Lan and 379 Da for MeLan). **(a)** Hydrolyzed and derivatized MeLan residues from modified CylL$_L$ (red line) and co-injected with derivatized DL-MeLan (2*S*,3*S*,6*R*) standard (blue line). **(b)** Hydrolyzed and derivatized MeLan residues from modified CylL$_L$ (red line) and co-injected with derivatized LL-MeLan (2*R*,3*R*,6*R*) standard (blue line). **(c)** Hydrolyzed and derivatized Lan residues from modified CylL$_L$ (red line) and co-injected with derivatized DD-Lan (2*S*,6*S*) standard (blue line). **(d)** Hydrolyzed and derivatized Lan residues from modified CylL$_L$ (red line) and co-injected with derivatized DL-Lan (*meso*) standard (blue line). **(e)** Hydrolyzed and derivatized Lan residues from modified CylL$_L$ (red line) and co-injected with derivatized LL-Lan (2*R*,6*R*) standard (blue line). Traces of derivatized Lan and MeLan residues from modified CylL$_L$ (red lines) used for overlay were adjusted to 70% intensity for clarity. A small amount of epimerization of the (Me)Lan occurs during acid hydrolysis of the peptides as has been reported previously (*22*).

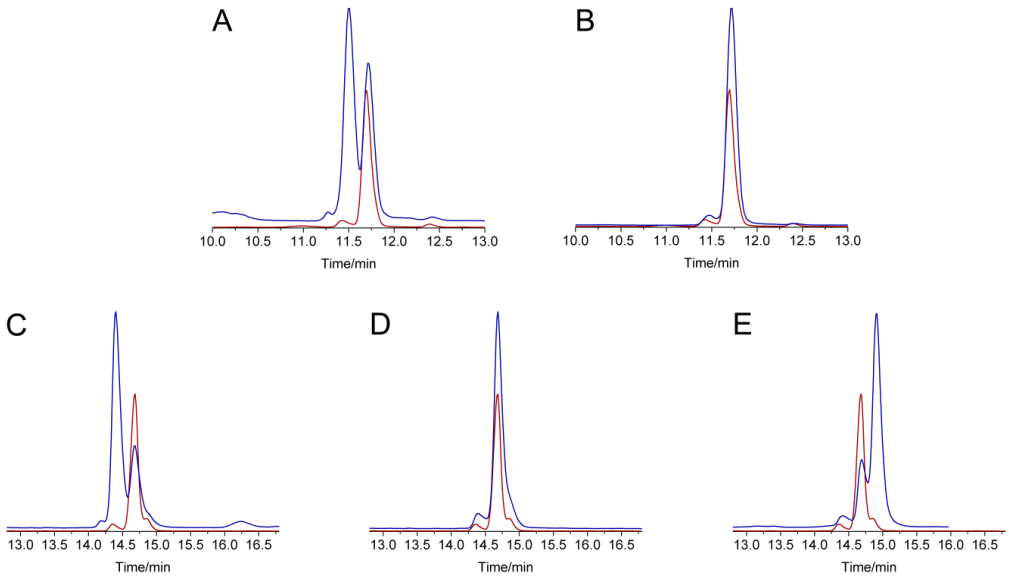

Figure 2.8 GC-MS traces for co-injections of synthetic, derivatized Lan/MeLan standards and hydrolyzed/derivatized Lan/MeLan residues obtained from *in vivo* modified CylL$_S$ (selected ion monitoring, SIM, at 365 Da for Lan and 379 Da for MeLan). **(a)** Hydrolyzed and derivatized MeLan residues from modified CylL$_S$ (red line) and co-injected with derivatized DL-MeLan standard (blue line). **(b)** Hydrolyzed and derivatized MeLan residues from modified CylL$_S$ (red line) and co-injected with derivatized LL-MeLan standard (blue line). **(c)** Hydrolyzed and derivatized Lan residues from modified CylL$_S$ (red line) and co-injected with derivatized DD-Lan standard (blue line). **(d)** Hydrolyzed and derivatized Lan residues from modified CylL$_S$ (red line) and co-injected with derivatized DL-Lan standard (blue line). **(e)** Hydrolyzed and derivatized Lan residues from modified CylL$_S$ (red line) and co-injected with derivatized LL-Lan standard (blue line). Traces of derivatized Lan and MeLan residues from modified CylL$_S$ (red lines) used for overlay were adjusted to 70% intensity for clarity. A small amount of epimerization of the (Me)Lan occurs during acid hydrolysis of the peptides as has been reported previously (*22*).

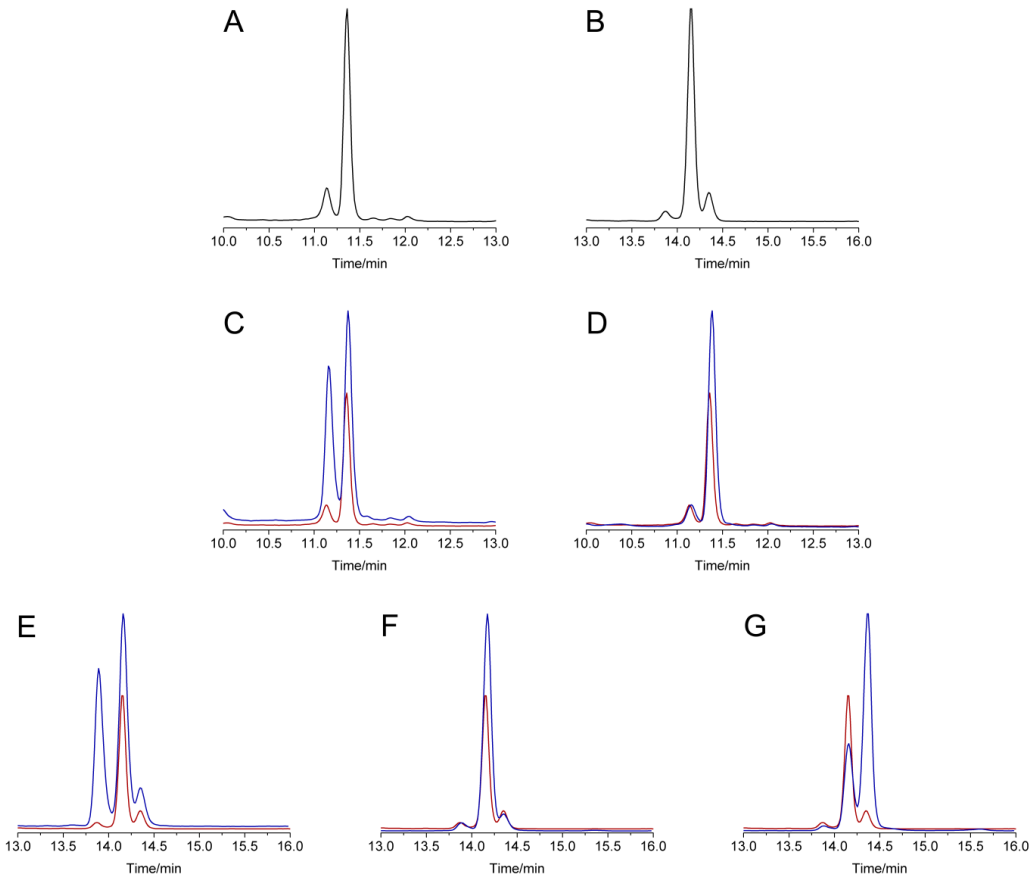

Figure 2.9 GC-MS traces for hydrolyzed/derivatized Lan/MeLan residues obtained from *in vivo* modified CylL$_L$ S14T and co-injections with derivatized Lan/MeLan standards (selected ion monitoring, SIM, at 365 Da for Lan and 379 Da for MeLan). **(a)** Hydrolyzed and derivatized MeLan residues from modified CylL$_L$ S14T. **(b)** Hydrolyzed and derivatized Lan residues from modified CylL$_L$ S14T. **(c)** Hydrolyzed and derivatized MeLan residues from modified CylL$_L$ S14T (red line) and co-injected with derivatized DL-MeLan standard (blue line). **(d)** Hydrolyzed and derivatized MeLan residues from modified CylL$_L$ S14T (red line) and co-injected with derivatized LL-MeLan standard (blue line). **(e)** Hydrolyzed and derivatized Lan residues from modified CylL$_L$ S14T (red line) and co-injected with derivatized DD-Lan standard (blue line). **(f)** Hydrolyzed and derivatized Lan residues from modified CylL$_L$ S14T (red line) and co-injected with derivatized DL-Lan standard (blue line). **(g)** Hydrolyzed and derivatized Lan residues from modified CylL$_L$ S14T (red line) and co-injected with derivatized LL-Lan standard (blue line). Traces of derivatized Lan and MeLan residues from modified CylL$_L$ S14T (red lines) used for overlay were adjusted to 70% intensity for clarity. A small amount of epimerization of the (Me)Lan occurs during acid hydrolysis of the peptides as has been reported previously (*22*).

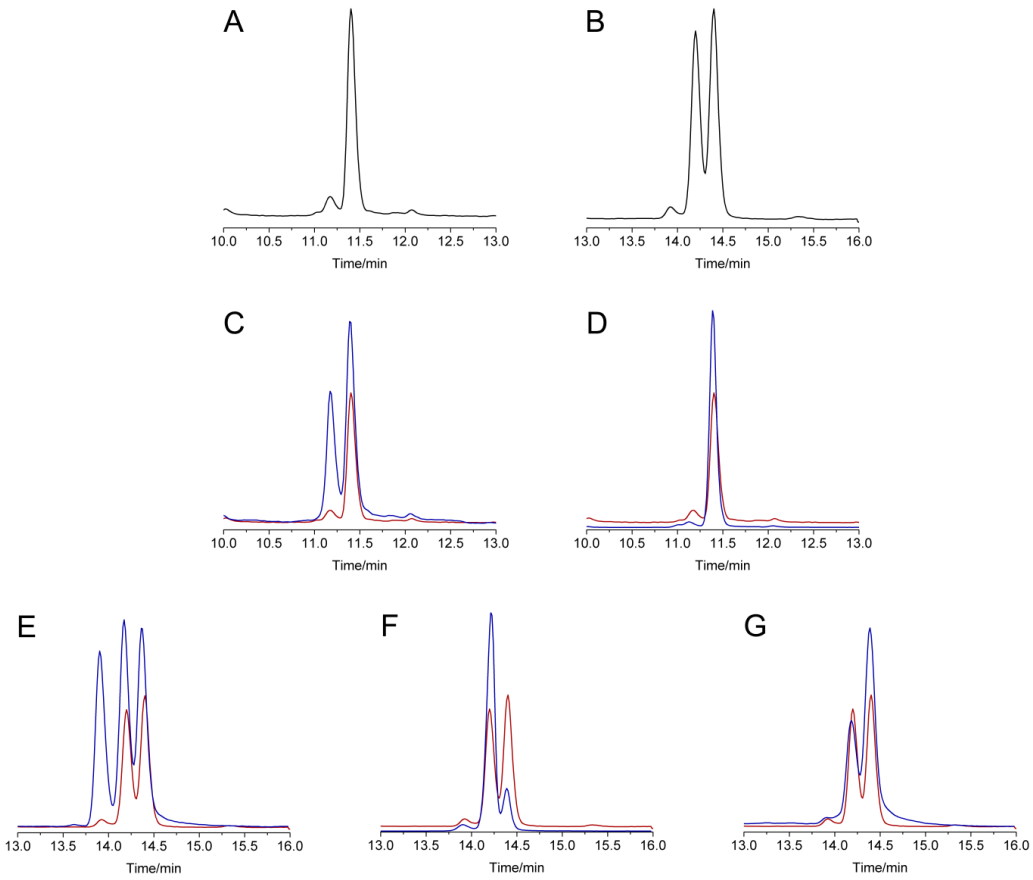

Figure 2.10 GC-MS traces for hydrolyzed/derivatized Lan/MeLan residues obtained from *in vivo* modified CylL$_L$ S15T and co-injections with derivatized Lan/MeLan standards (selected ion monitoring, SIM, at 365 Da for Lan and 379 Da for MeLan). **(a)** Hydrolyzed and derivatized MeLan residues from modified CylL$_L$ S15T. **(b)** Hydrolyzed and derivatized Lan residues from modified CylL$_L$ S15T. **(c)** Hydrolyzed and derivatized MeLan residues from modified CylL$_L$ S15T (red line) and co-injected with derivatized DL-MeLan standard (blue line). **(d)** Hydrolyzed and derivatized MeLan residues from modified CylL$_L$ S15T (red line) and co-injected with derivatized LL-MeLan standard (blue line). **(e)** Hydrolyzed and derivatized Lan residues from modified CylL$_L$ S15T (red line) and co-injected with derivatized DD-Lan standard (blue line). **(f)** Hydrolyzed and derivatized Lan residues from modified CylL$_L$ S15T (red line) and co-injected with derivatized DL-Lan standard (blue line). **(g)** Hydrolyzed and derivatized Lan residues from modified CylL$_L$ S15T (red line) and co-injected with derivatized LL-Lan standard (blue line). Traces of derivatized Lan and MeLan residues from modified CylL$_L$ S15T (red lines) used for overlay were adjusted to 70% intensity for clarity. A small amount of epimerization of the (Me)Lan occurs during acid hydrolysis of the peptides as has been reported previously (*22*).

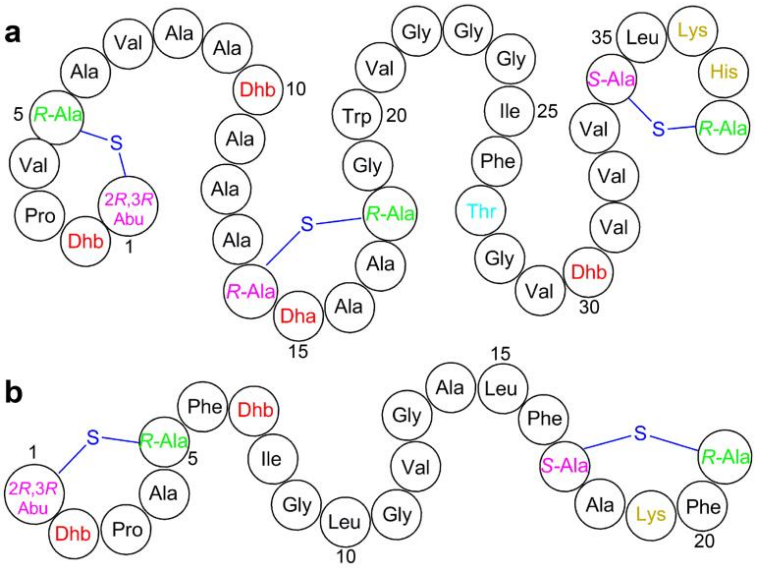

Figure 2.11 Structure of cytolysin. (a) $CylL_L$". **(b)** $CylL_S$". Abu, α-aminobutyric acid. Overall, these peptides are highly hydrophobic except for the positively charged N-termini and residues (Lys and His) near the C-terminus (marked in dark yellow). Production of cytolysin in *E. coli* allows investigation of the importance of these residues for the unique hemolytic activity of this lantibiotic.

The data clearly demonstrate that LL-MeLan can be formed in the same peptide that also contains Lan with the more common DL-stereochemistry. This highly unusual observation appears to be at odds with the typical high stereochemical fidelity of enzymes. Interestingly, all (Me)Lan residues that exhibited the LL configuration are formed from a Dhx-Dhx-Xxx-Xxx-Cys motif in which Dhx represents Dha or Dhb, and Xxx represents amino acids other than Dha, Dhb and Cys (**Figure 2.12)**. To probe whether this outcome extends to other naturally occurring lantibiotics, I performed searches for the Dhx-Dhx-Xxx-Xxx-Cys motif in documented lantibiotics (**Figure 2.13a**). Several such motifs were found (*17*, *18*, *23*, *24*) allowing experimental probing of the hypothesis. Halβ, one subunit of the two-component lantibiotic haloduracin produced by *Bacillus halodurans* C-125 (*18*), was chosen as an example. Halβ was obtained by co-expression of the precursor peptide HalA2 with its lanthionine synthetase HalM2 in *E. coli* (*16*). To investigate the stereochemistry of the (Me)Lan in the A ring, the HalM2-modified HalA2 was hydrolyzed, and the amino acids were derivatized and analyzed by chiral GC-MS. The data show clearly that the product contained a 1:2 ratio of LL:DL MeLan and only DL-Lan (**Figures 2.13b** and **2.14**), consistent with the hypothesis because only one of the three MeLan structures in Halβ is generated from a Dhx-Dhx-Xxx-Xxx-Cys motif (the A ring). The assignment of the LL-MeLan to this A ring

of Halβ was supported by GC-MS analysis of the HalM2-modified mutant HalA2-C5A, for which the peak for LL-MeLan was reduced to a basal level (**Figure 2.15**). These results therefore show that LL-MeLan residues are present in other lantibiotics, and that their formation is not a special property of CylM. The LL stereochemistry is also not caused by expression in *E. coli* because the same results were obtained for Halβ isolated from *B. halodurans* (**Figure 2.16**). Analysis of the databases suggest that other lantibiotics that have been previously reported but for which the stereochemistry has not been determined are also likely to contain LL-(Me)Lan residues (**Figure 2.13a**). In turn, these studies also indicate that the sequence of the substrate peptide can determine the stereoselectivity of (Me)Lan formation, which was unanticipated prior to this study and has important implications for lantibiotic engineering. Whether the determinant is the sequence itself or the position of the sequence within the precursor peptide, or both, required additional studies, some of which will be presented in chapter 3.

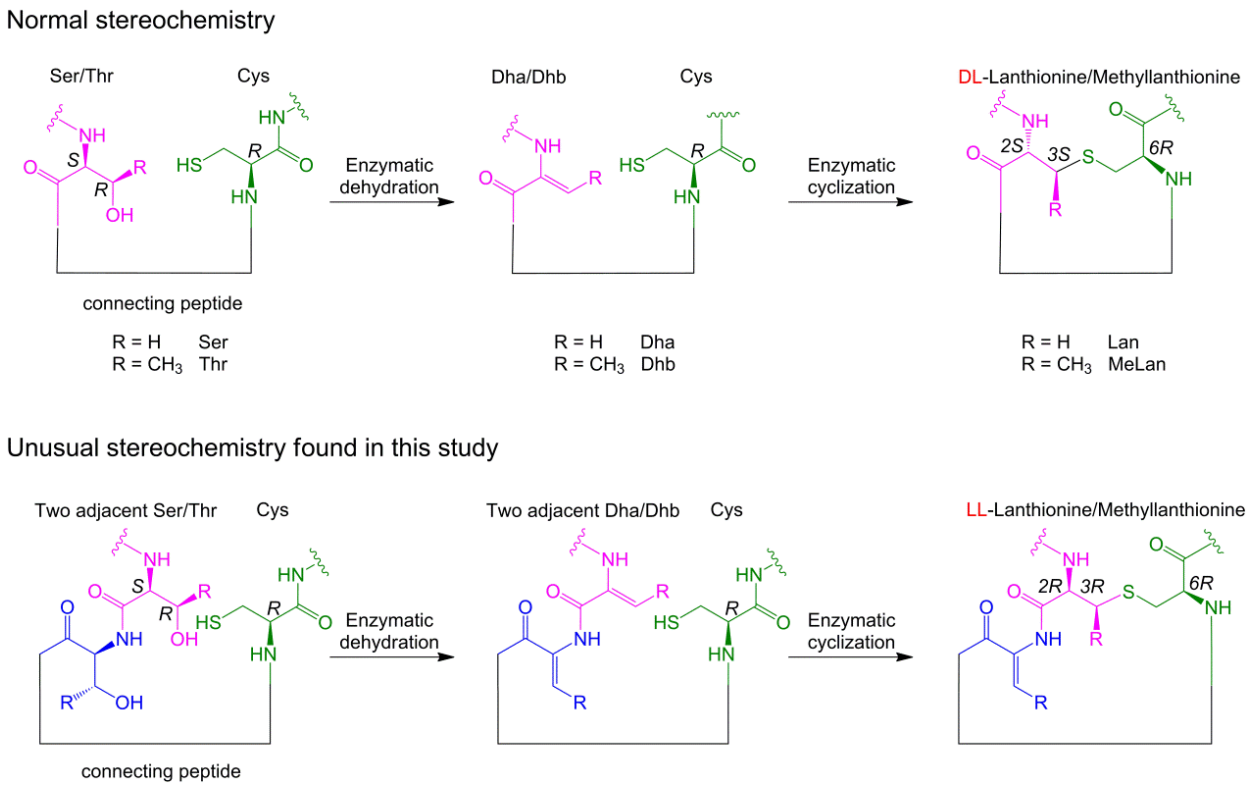

Figure 2.12 Enzyme catalyzed formation of the canonical DL-(Me)Lan (top) and the unusual LL-(Me)Lan (bottom) from the Dhx-Dhx-Xxx-Xxx-Cys motif.

Purified CylL$_L$" and CylL$_S$" were tested for their antimicrobial activity against *Lactococcus lactis* HP. Application of CylL$_L$" and CylL$_S$" separately did not result in growth inhibition, but

when applied together, a strong inhibition zone was observed (**Figure 2.13c**). Furthermore, when the two compounds were placed a short distance from each other, a synergistic zone of growth inhibition was only observed between the two spotted samples (**Figure 2.13c**). The hemolytic activity of cytolysin was investigated using rabbit red blood cells (**Figure 2.13d**). CylL$_L$" and CylL$_S$" together lead to maximal cell lysis in 90 min at a concentration as low as 200 nM, whereas either CylL$_L$" or CylL$_S$" alone induced minimal cell lysis at a concentration of 800 nM. The synergistic anti-microbial and hemolytic activities of the CylL$_L$" and CylL$_S$" prepared in *E. coli*, along with the MS characterization, suggests that the cytolysin produced in this work has the same structure as that obtained from the producing strains in previous studies (*1*, *25*).

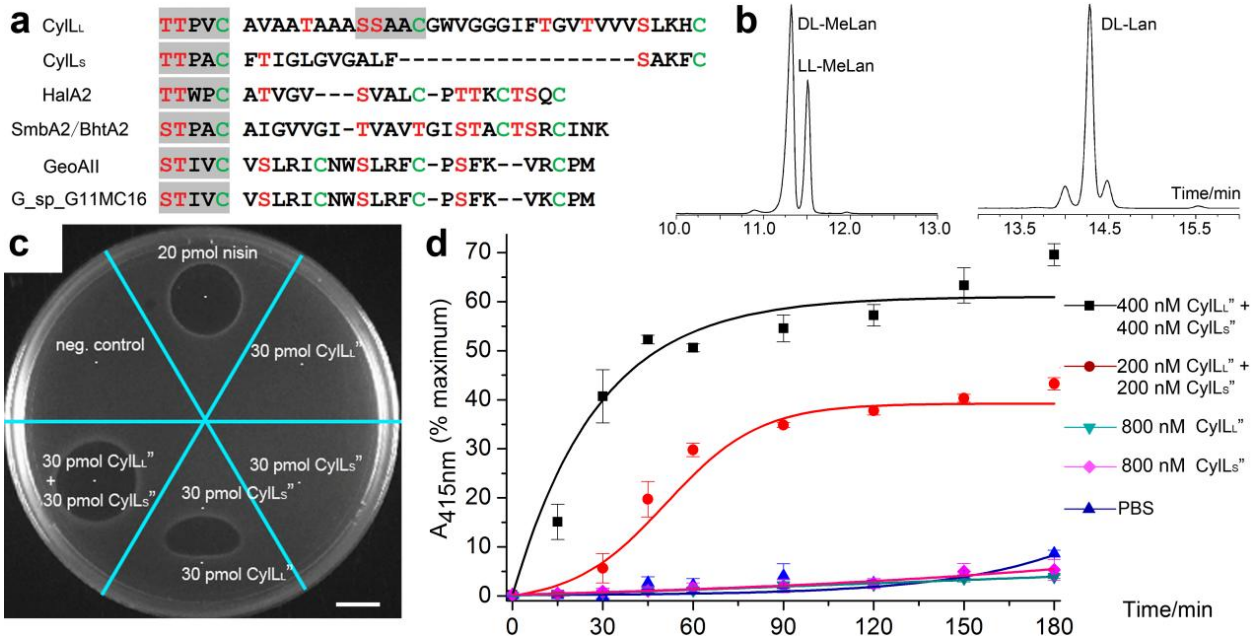

Figure 2.13 Alignment of 7 lantibiotic core peptides, stereochemical analysis of Halβ, and bioactivity of CylL$_L$" and CylL$_S$". (a) Alignment of lantibiotic core peptides that are transformed to intermediates with the Dhx-Dhx-Xxx-Xxx-Cys motifs that are highlighted in grey. The amino acid sequences of SmbA2 and BhtA2 are identical, but it is unclear whether they are modified into the same structure. **(b)** GC-MS traces of hydrolyzed and derivatized MeLan and Lan residues from HalM2-modified HalA2. **(c)** Antimicrobial activity assay of CylL$_L$" and CylL$_S$" against *L. lactis* HP. The amounts of compounds applied are indicated (scale bar 1 cm). **(d)** Hemolytic assay of CylL$_L$" and CylL$_S$" against rabbit red blood cells. At each time point, the hemoglobin released from the lysed cells was quantified in triplicate by measuring the absorbance of the supernatant at 415 nm. Error bars indicate the standard deviation of three separate experiments. The lytic curve was fit with a Growth-Sigmoidal-Dose Response function, which does not necessarily correlates to the kinetic properties of the lytic reaction.

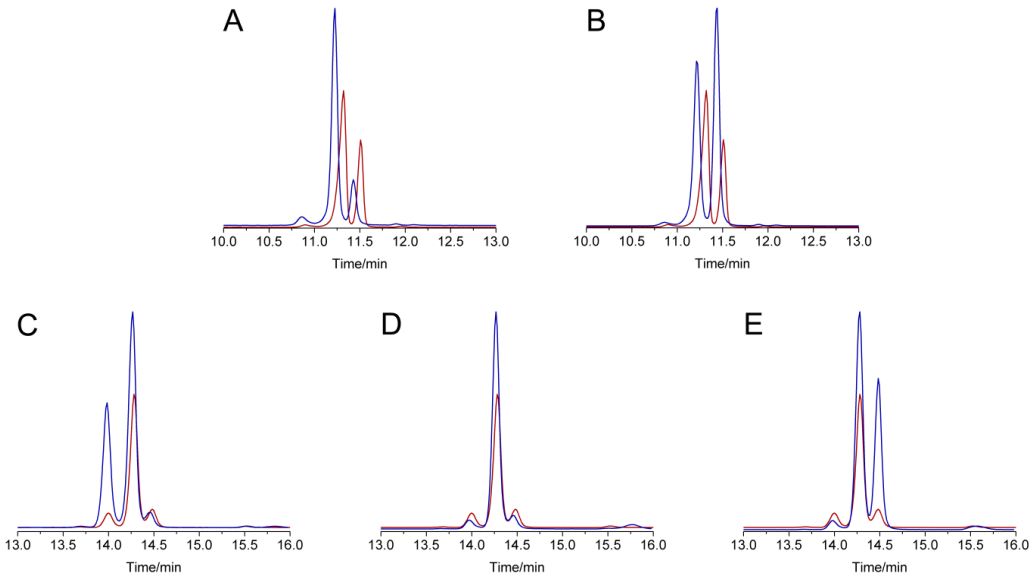

Figure 2.14 GC-MS traces for co-injections of synthetic, derivatized Lan/MeLan standards and hydrolyzed/derivatized Lan/MeLan residues obtained from *in vivo* modified HalA2Xa (selected ion monitoring, SIM, at 365 Da for Lan and 379 Da for MeLan). **(a)** Hydrolyzed and derivatized MeLan residues from modified HalA2Xa (red line) and co-injected with derivatized DL-MeLan standard (blue line). **(b)** Hydrolyzed and derivatized MeLan residues from modified HalA2Xa (red line) and co-injected with derivatized LL-MeLan standard (blue line). **(c)** Hydrolyzed and derivatized Lan residues from modified HalA2Xa (red line) and co-injected with derivatized DD-Lan standard (blue line). **(d)** Hydrolyzed and derivatized Lan residues from modified HalA2Xa (red line) and co-injected with derivatized DL-Lan standard (blue line). **(e)** Hydrolyzed and derivatized Lan residues from modified HalA2Xa (red line) and co-injected with derivatized LL-Lan standard (blue line). Traces of derivatized Lan and MeLan residues from modified HalA2Xa (red lines) used for overlay were adjusted to 70% intensity for clarity. HalA2Xa is a derivative of HalA2 in which a Factor Xa cleavage site has been introduced between the leader and core peptides (*26*). A small amount of epimerization of the (Me)Lan occurs during acid hydrolysis of the peptides as has been reported previously (*22*).

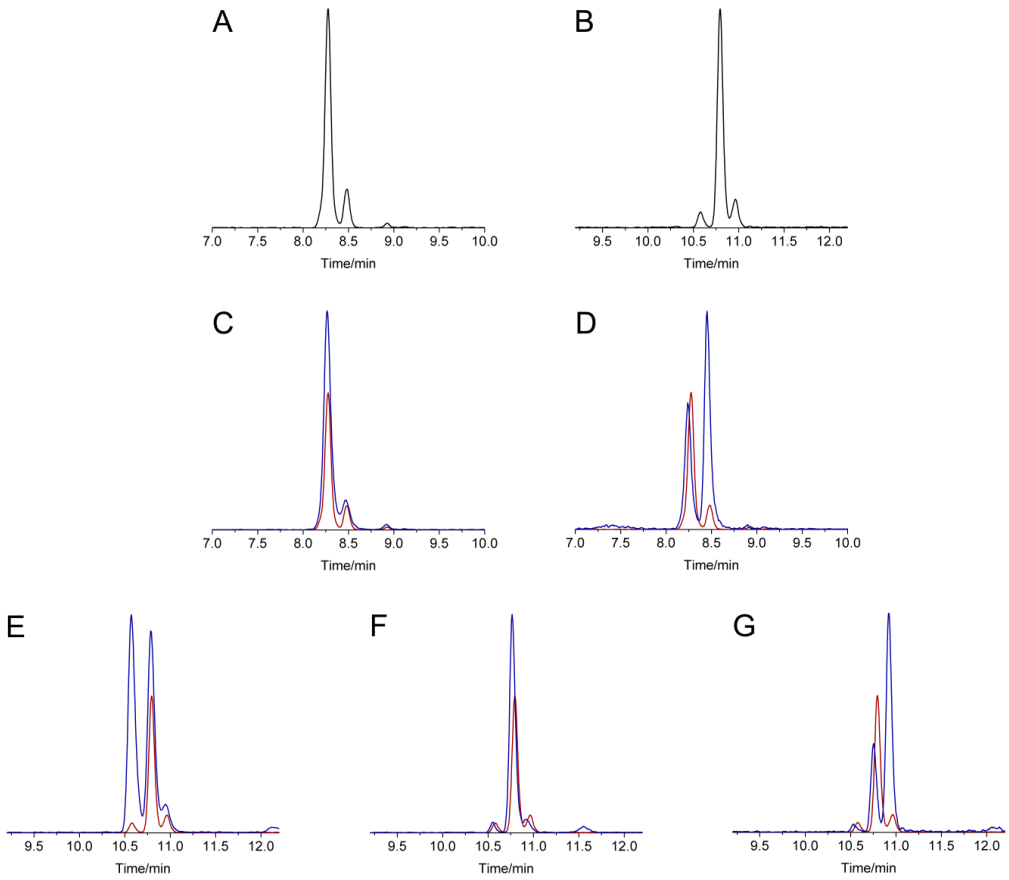

Figure 2.15 GC-MS traces for hydrolyzed/derivatized Lan/MeLan residues obtained from *in vivo* modified HalA2Xa C5A and their co-injections with derivatized Lan/MeLan standards (selected ion monitoring, SIM, at 365 Da for Lan and 379 Da for MeLan). **(a)** Hydrolyzed and derivatized MeLan residues from modified HalA2Xa C5A. **(b)** Hydrolyzed and derivatized Lan residues from modified HalA2Xa C5A. **(c)** Hydrolyzed and derivatized MeLan residues from modified HalA2Xa C5A (red line) and co-injected with derivatized DL-MeLan standard (blue line). **(d)** Hydrolyzed and derivatized MeLan residues from modified HalA2Xa C5A (red line) and co-injected with derivatized LL-MeLan standard (blue line). **(e)** Hydrolyzed and derivatized Lan residues from modified HalA2Xa C5A (red line) and co-injected with derivatized DD-Lan standard (blue line). **(f)** Hydrolyzed and derivatized Lan residues from modified HalA2Xa C5A (red line) and co-injected with derivatized DL-Lan standard (blue line). **(g)** Hydrolyzed and derivatized Lan residues from modified HalA2Xa C5A (red line) and co-injected with derivatized LL-Lan standard (blue line). Traces of derivatized Lan and MeLan residues from modified HalA2Xa C5A (red lines) used for overlay were adjusted to 70% intensity for clarity. A small amount of epimerization of the (Me)Lan occurs during acid hydrolysis of the peptides as has been reported previously (*22*).

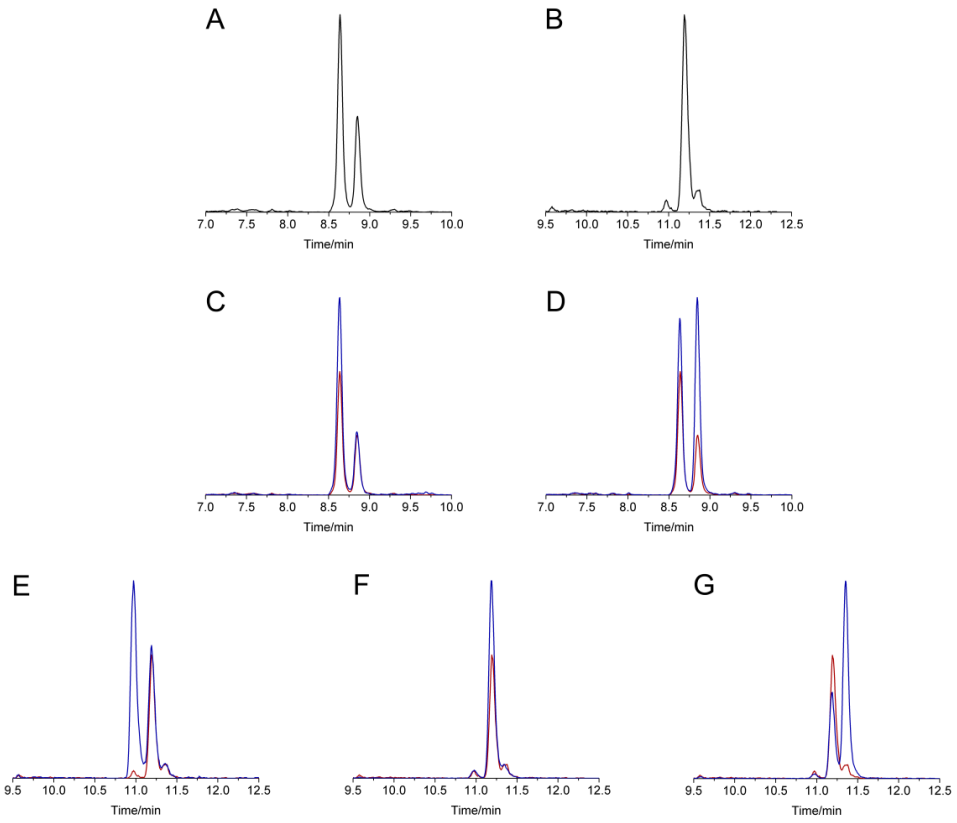

Figure 2.16 GC-MS traces for hydrolyzed/derivatized Halβ isolated from *B. halodurans* C-125 and co-injections with derivatized Lan/MeLan standards (selected ion monitoring, SIM, at 365 Da for Lan and 379 Da for MeLan). **(a)** Hydrolyzed and derivatized MeLan residues from modified Halβ. **(b)** Hydrolyzed and derivatized Lan residues from Halβ. **(c)** Hydrolyzed and derivatized MeLan residues from Halβ (red line) and co-injected with derivatized DL-MeLan standard (blue line). **(d)** Hydrolyzed and derivatized MeLan residues from Halβ (red line) and co-injected with derivatized LL-MeLan standard (blue line). **(e)** Hydrolyzed and derivatized Lan residues from Halβ (red line) and co-injected with derivatized DD-Lan standard (blue line). **(f)** Hydrolyzed and derivatized Lan residues from Halβ (red line) and co-injected with derivatized DL-Lan standard (blue line). **(g)** Hydrolyzed and derivatized Lan residues from Halβ (red line) and co-injected with derivatized LL-Lan standard (blue line). Traces of derivatized Lan and MeLan residues from Halβ (red lines) used for overlay were adjusted to 70% intensity for clarity. A small amount of epimerization of the (Me)Lan occurs during acid hydrolysis of the peptides as has been reported previously (*22*).

As a summary of this chapter, I have developed a heterologous production system for the enterococcal cytolysin that is strongly linked to virulence, and unambiguously characterized its structure. Three non-overlapping five-amino acid rings are present in $CylL_L$" and two such rings are present in $CylL_S$". The MeLan residues in the A rings of both $CylL_L$" and $CylL_S$" and the Lan

residue in the B ring of $CylL_L$" exhibit the LL configuration that is different from all previously characterized lantibiotics. LL-(Me)Lan residues are predicted to also be present in other lantibiotics, and this prediction was confirmed for the A ring of Halβ. These results emphasize the importance of determining the stereochemistry of lantibiotics, which has thus far only been performed for a handful of family members (*19-21, 27*). With a good supply route to the enterococcal cytolysin as well as analogs, the mechanism of its virulence activity can now be further investigated (*28*).

2.3 Methods

General methods

Genes encoding CylM, $CylL_L$ and $CylL_S$ were synthesized by GeneArt (Invitrogen) with codon usage optimized for *E. coli* expression (for DNA sequences, see **Table 2.1**). All polymerase chain reactions (PCR) were carried out on a C1000 thermal cycler (Bio-Rad). DNA sequencing was performed by ACGT, Inc. Preparative HPLC was performed using a Waters Delta 600 instrument equipped with appropriate columns. Absorbance of rabbit hemoglobin in 96-well plates was measured with a Synergy H4 Microplate Reader (BioTek). GC-MS analysis was performed at the Roy J. Carver Metabolomics Center (UIUC) or Mass Spectrometry Laboratory (School of Chemical Sciences, UIUC) on an Agilent 7890 gas chromatograph (Agilent). LC-ESI-Q/TOF MS analyses were conducted using a Synapt MS system equipped with an Acquity UPLC (Waters). High resolution electrospray ionization (HR-ESI) MS was performed on a Micromass Q-Tof Ultima instrument (Waters).

Table 2.1 Synthetic DNA sequences for *cylM, cylL*$_L$ and *cylL*$_S$ with codon usage optimized for *E. coli* expression.

cylM
ATGGAAGATAATCTGATTAATGTGCTGAGCATTAATGAACGTTGCTTTCTGCTGAAACAGAGCGGCAA
AGAAAAATATGATATTAAAAATCTGCAGGCCTGGAAAGAACGTAAAAGCGTTCTGAAACAGGATGAT
CTGGATTATCTGATTAAATATAAATATGAAAGCCTGGATAATTTTGGCCTGGGTATTACCCCGATTGAA
AATTTTCCGGATAAAGAAGTGGCCATTCAGTATATTAAAGATCAGAGTTGGTATATTTTTTTTGAAAGC
ATTCTGGATAGCTATAATGATAGCGAAGAAAAACTGCTGGAAGTTGATGCAAGCTATCCGTTTCGTTA
TTTTCTGCAGTATGCACGTCTGTTTCTGCTGGATCTGAATAGCGAACTGAATATTTGCACCAAAGAATT
TATTATTAATCTGCTGGAAACCCTGACCCAGGAACTGATTCATCTGACCAGCAAAACCCTGGTTCTGG
ATCTGCATACCTTTAAAAAAAATGAACCGCTGAAAGGCAATGATAGCAGCAAACGCTTTATTTATTAT

Table 2.1 (cont.)

CTGAAAAAACGCTTTAATAGCAAAAAAGATATTATTGCCTTTTATACCTGCTATCCGGAACTGATGCGT
ATTACCGTTGTTCGTATGCGCTATTTCCTGGATAACACCAAACAAATGCTGATTCGTGTTACCGAAGAT
CTGCCGAGCATTCAGAATTGCTTTAATATTCAGAGCAGTGAACTGAATAGCATTAGCGAAAGCCAGGG
TGATAGCCATAGCCGTGGTAAAACCGTTAGCACCCTGACCTTTAGTGATGGTAAAAAAATTGTGTATA
AACCGAAAATTAATAGCGAAAATAAACTGCGCGATTTTTTTGAATTTCTGAATAAAGAACTGGAAGCC
GATATTTATATTGTGAAAAAAGTGACCCGCAATACCTATTTTTATGAAGAATATATCGATAACATTGA
AATCAATAACATCGAAGAAGTGAAAAAATATTACGAACGCTATGGCAAACTGATTGGCATTGCCTTTC
TGTTTAATGTTACCGATCTGCATTATGAAAACATCATTGCCCATGGCGAATATCCGGTGATTATTGATA
ATGAAACCTTTTTTCAGCAGAATATTCCGATTGAATTTGGTAATAGCGCAACCGTTGATGCCAAATACA
AATATCTGGATAGCATTATGGTGACCGGTCTGGTTCCGTATCTGGCAATGAAAGATAAAAGCGATAGC
AAAGATGAAGGCGTTAATCTGAGCGCACTGAATTTTAAAGAACAGAGCGTGCCGTTTAAAATTCTGAA
AATTAAAAATACCTTTACCGATGAAATGCGCTTTGAATATCAGACCCATATTATGGATACCGCAAAAA
ATACCCCGATTATGAATAATGAAAAAATCAGCTTTATCAGCTATGAAAAATATATTGTGACCGGCATG
AAAAGCATTCTGATGAAAGCCAAAGATAGCAAAAAAAAAATTCTGGCCTATATTAATAATAATCTGCA
GAATCTGATTGTGCGCAATGTTATTCGTCCGACCCAGCGTTATGCAGATATGCTGGAATTTAGCTATCA
TCCGAATTGCTTTAGCAATGCCATTGAACGTGAAAAAGTGCTGCATAATATGTGGGCCTATCCGTATA
AAAATAAAAAAGTGGTGCATTATGAATTTTCAGATCTGATTGATGGCGATATTCCGATTTTTTATAATA
ATATTAGCAAAACCAGCCTGATTGCCAGTGATGGTTGTCTGGTTGAAGATTTTTATCAGGAAAGTGCC
CTGAATCGCTGCCTGAATAAAATTAATGATCTGTGTGATGAAGATATTAGCATTCAGACCGTGTGGCT
GGAAATTGCACTGAATATTTATAACCCGTACAAATATATCAATGATCTGAAAAACCAGAATAGCAACA
AATATATTTATACCGGTCTGGAACTGAATGGCAAAATTATTCAGGCCTGCCAGAAAATTGAAAAAAAA
ATCTTTAAACGTGCCATCTTTAACAAAAAAACCAATACCGTGAATTGGATTGATATTAAACTGGATCA
GGATTGGAATGTGGGCATTCTGAATAATAATATGTATGATGGCCTGCCTGGTATTTTTATTTTTTATGT
GGCCCTGAAATATATTACCAAAAACCATAAATATGATTATGTGATCGAATGCATTAAAAATAGCATTT
ATACCATTCCGAGCGAAGATATTCTGAGCGCATTTTTTGGTAAAGGCAGCCTGATTTATCCGCTGCTGG
TTGATTATCGCCTGAATAATGATATTAATAGCCTGAATGTGGCCGTGGAAATTGCCGATATGCTGATTG
AAAAAAAACCGATTAATAATGGCGAACTGAAAAATGATTGGATTCATGGCCATAATAGCATTATTAAA
GTGCTGCTGCTGCTGAGCGAAATTACCGAAGATGAAAAATATCGCAAATTTAGCCTGGAAATTTTTGA
AAAACTGAGCGAAGAACCGTATTTTAATTTTCGTGGTTTTGGCCATGGCATTTATAGCTATGTTCATCT
GCTGAGCAAATTTAATCGCATTGATAAAGCCAATAGCCTGCTGCATAAAATTAAAGAAAGCTATTTTG
AAGAAGAACCGAAAAATAATTCCTGGTGTAAAGGCACCGTTGGTGAACTGCTGGCAACCATTGAACT
GTATGATGATAATATTAGTAACATCGATATTAACAAAACCATTGCCTATAAAAATAAAGATTGCCTGT
GCCATGGCAATGCAGGCACCCTGGAAGGTCTGATTCAGCTGGCAAAAAAAGATCCGGAAACCTATCA
GTATAAAAAAAATAAACTGATCAGCTATATGCTGAATGATTTTGAAAAAAATAATACCCTGAAAGTGG
CAGGCAGCGAATATCTGGAAAGCCTGGGTTTTTTTGTTGGTATTAGCGGTGTTGGTTATGAACTGCTGC
GTAATCTGGATAGCGAAATTCCGAATGCACTGCTGTTTGAACTGTAATAA

cylL$_L$

ATGGAAAATCTGAGCGTTGTTCCGAGCTTTGAAGAACTGAGCGTTGAAGAAATGGAAGCAATTCAGG
GTAGCGGTGATGTTCAGGCAGAAACCACACCGGTTTGTGCAGTTGCAGCAACCGCAGCAGCAAGCAG
CGCAGCATGTGGTTGGGTTGGTGGTGGTATTTTTACCGGTGTTACCGTTGTTGTTAGCCTGAAACATTG
CTAATAA

cylL$_S$

ATGCTGAATAAAGAAAATCAGGAAAATTATTATAGCAATAAACTGGAACTGGTGGGTCCGAGCTTTGA

AGAACTGAGCCTGGAAGAAATGGAAGCAATTCAGGGTAGCGGTGATGTTCAGGCAGAAACCACACCG
GCATGTTTTACCATTGGTCTGGGTGTTGGTGCACTGTTTAGCGCAAAATTTTGCTAATAA

Materials

All oligonucleotides were purchased from Integrated DNA Technologies and used as received. Restriction endonucleases, DNA polymerases, and T4 DNA ligase were from New England Biolabs. Media components were obtained from Difco Laboratories. Trypsin was purchased from Worthington Biochemical Corporation and other endoproteinases were ordered from Roche Biosciences. Unless specified otherwise, chemicals were purchased from Sigma Aldrich or Fisher Scientific. Defibrinated rabbit blood was ordered from Hemostat Laboratories and used within 10 days of receipt.

Strains and plasmids

The indicator strain, *Lactococcus lactis* HP, and the haloduracin producer strain *Bacillus halodurans* C-125 was obtained from American Type Culture Collection. *E. coli* DH5α and *E. coli* BL21 (DE3) cells were used as host for cloning and plasmid propagation, and host for expression, respectively. Co-expression vector pRSFDuet-1 was obtained from Novagen.

Construction of pRSFDuet-1 derivatives for co-expression of CylM and $CylL_L$/$CylL_S$

The gene encoding CylM was first cloned into the multiple cloning site 2 of the pRSFDuet-1 vector (without His-tag sequence) using *Nde*I and *Kpn*I restriction sites to generate pRSFDuet-1/CylM-2. The $cylL_L$ and $cylL_S$ genes were then cloned into the multiple cloning site 1 (MCS1) of pRSFDuet-1/CylM-2 (with His-tag sequence) using *EcoR*I and *Not*I restriction sites to generate pRSFDuet-1/$CylL_L$/CylM-2 and pRSFDuet-1/$CylL_S$/CylM-2 plasmids, respectively. Primer sequences are shown in **Table 2.2**.

Table 2.2 Primer sequences for cloning of *cylM*, *cylL$_L$*, *cylL$_S$*, *halA2* and their mutants. Mutation codons are underlined.

Primer Name	Primer Sequence (5'-3')
CylM_NdeI_FP_Duet	AAAAA CATATG GAAGATA ATCTGATTAA T
CylM_KpnI_RP_Duet	AAAAA GGTACC TTACAGT TCAAACAGCA G
CylL$_L$_EcoRI_FP_Duet	AAAAA GAATTCG GAAAATC TGAGCGTTGT T
CylL$_L$_NotI_RP_Duet	AAAAA GCGGCCGC TTAGCAA TGTTTCAGGC T
CylL$_S$_EcoRI_FP_Duet	AAAAA GAATTCG CTGAATA AAGAAAATCA G
CylL$_S$_NotI_RP_Duet	AAAAA GCGGCCGC TTAGCAA AATTTTGCGC T
CylL$_L$_E-1K_FP	ACC ACA CCG GTT TGT GCA GTT
CylL$_L$_E-1K_RP	ACA AAC CGG TGT GGT TTT TGC CTG AAC ATC ACC
CylL$_S$_E-1K_FP	ACC ACA CCG GCA TGT TTT ACC
CylL$_S$_E-1K_RP	ACA TGC CGG TGT GGT TTT TGC CTG AAC ATC ACC
CylL$_L$_S14T_FP	GCA ACC GCA GCA GCA ACC AGC GCA GCA TGT GGT
CylL$_L$_S15T_FP	GCA ACC GCA GCA GCA AGC ACC GCA GCA TGT GGT
CylL$_L$_S14/15T_RP	TGC TGC TGC GGT TGC TGC AAC
HalA2_BamHI_FP_Duet	AAAAA GGATCC G ATGGTAAATT CAAAAGATTT
HalA2_HindIII_RP_Duet	AAAAA AAGCTT TTAGCACTGG CTTGTACACT

Construction of pRSFDuet-1 derivatives for co-expression of CylM and CylL$_L$/CylL$_S$-E−1K, CylL$_L$-S14T, and CylL$_L$-S15T

Site-directed mutagenesis of pRSFDuet-1/CylL$_L$/CylM-2 and pRSFDuet-1/CylL$_S$/CylM-2 constructs was performed by multi-step overlap extension PCR. Negative numbers are used for amino acids in the leader peptide counting backwards from the leader peptide cleavage site. First, the amplification of the 5' part of *cylL$_L$/cylL$_S$* was carried out by 30 cycles of denaturing (95 ℃ for 10 s), annealing (55 ℃ for 30 s), and extending (72 ℃ for 15 s) using the *cylL$_L$/cylL$_S$ Eco*RI forward primer (FP) and an appropriate mutant reverse primer (**Table 2.2**) to generate a forward megaprimer (FMP). The PCR mixture included *Phusion* DNA polymerase (Finnzymes; 0.04 U/μL), dNTP (2 mM each), primers (1 μM each), and appropriate buffers. In parallel, a PCR reaction using an appropriate mutant forward primer (**Table 2.2**) and the *cylL$_L$/cylL$_S$ Not*I reverse primer (RP) was also performed to produce 3' fragments (termed reverse megaprimer, RMP). The 5' FMP fragment and the 3' RMP fragment were purified by 2% agarose gel and combined in equal amounts (by mol, approximately 20 ng each for a 50 μL PCR reaction) and amplified using the same PCR conditions with *cylL$_L$/cylL$_S$ Eco*RI FP and *cylL$_L$/cylL$_S$ Not*I RP. The resulting PCR

products were then cloned into the MCS1 of pRSFDuet-1/CylM-2 to generate pRSFDuet-1/CylL$_L$-E−1K/CylM-2, pRSFDuet-1/CylL$_S$-E−1K/CylM-2, pRSFDuet-1/CylL$_L$-S14T/CylM-2 and pRSFDuet-1/CylL$_L$-S15T/CylM-2.

Construction of pRSFDuet-1 derivatives for co-expression of HalM2 and HalA2-C5A

The gene encoding HalA2-C5A was amplified by PCR from a pET15b/HalA2Xa-C5A vector reported previously (*26*), and cloned into the multiple cloning site 1 of pRSFDuet-1/HalM2-2 (*16*) using *BamH*I and *Hind*III restriction sites to generate the pRSFDuet-1/HalA2-C5A/HalM2-2 plasmid. Primer sequences used are shown in **Table 2.2**.

Expression and purification of modified His$_6$-CylL$_L$ and His$_6$-CylL$_S$ peptides

E. coli BL21 (DE3) cells were transformed with the pRSFDuet-1/CylM-2-CylL$_L$ or pRSFDuet-1/ CylL$_S$/CylM-2 plasmids and plated on LB plates containing 50 mg/L kanamycin. A single colony was picked and grown in 50 mL of LB with kanamycin at 37 ^{o}C for 12 h and the resulting culture was used to inoculate 3 L of LB. Cells were cultured at 37 ^{o}C until the OD at 600 nm reached 0.5, cooled and IPTG was added to a final concentration of 0.1 mM. Cells were cultured at 18 ^{o}C for another 18 h before harvesting. The cell pellet was resuspended at room temperature in LanA start buffer (20 mM NaH_2PO_4, pH 7.5 at 25 ^{o}C, 500 mM NaCl, 0.5 mM imidazole, 20% glycerol) and lysed by sonication. The sample was centrifuged at 23,700 ×*g* for 30 min and the supernatant was kept. The pellet was then resuspended in LanA buffer 1 (6 M guanidine hydrochloride, 20 mM NaH_2PO_4, pH 7.5 at 25 ^{o}C, 500 mM NaCl, 0.5 mM imidazole) and sonicated again. The insoluble portion was removed by centrifugation at 23,700 ×*g* for 30 min and the soluble portion was kept. Both soluble portions were passed through 0.45-µm syringe filters and the His-tagged modified peptides were purified by immobilized metal affinity chromatography (IMAC) as previously described (*29*). The eluted fractions were desalted by preparative reversed phase (RP) HPLC using a Waters Delta-pak C4 column (15 µm; 300 Å; 25 x 100 mm). The desalted peptides were lyophilized and stored at −20 ^{o}C for future use.

Protease cleavage and purification of core peptides

Modified $CylL_L$ and $CylL_S$ precursor peptides were dissolved in H_2O to a concentration of 3 mg/mL. To 85 µL of peptide solution, 10 µL of 500 mM HEPES buffer (pH 7.5) was added, followed by 5 µL of 0.02 mg/mL AspN protease (for wild type peptides) or 5 mg/mL trypsin (for the E−1K mutants). The protease cleavage reaction mixtures were kept at 25 ^{o}C (for wild type peptides) or 37 ^{o}C (for the E−1K mutants) for 6 to 16 h, respectively, and were then quenched with 0.5% TFA. The generation of the desired products was confirmed by MALDI-TOF MS and the products were purified by reversed phase (RP) HPLC using a Jupiter proteo C12 column (5 µm; 90 Å; 250 x 4.6 mm; Phenomenex). Solvents for RP-HPLC were solvent A (0.1% TFA in water) and solvent B (0.086% TFA in 80% acetonitrile/20% water). The desired core peptides eluted from the column between 50 and 70% solvent B and pure fractions were collected and lyophilized for further usage. Large scale purification was conducted by RP HPLC using a Jupiter proteo C12 semi-preparative column (5 µm; 90 Å; 250 x 10.0 mm).

LC-ESI-Q/TOF MS and MSMS analyses

A 5 µL volume of sample (approximately 100 µg/mL) was injected on a Waters Acquity UPLC system equipped with a BEH C8 column (1.7 um, 100 mm x 1.0 mm; Waters). The column was pre-equilibrated in aqueous solvent. The solvents used for LC were: solvent A = 0.1% formic acid in water and solvent B = 0.1% formic acid in methanol. A solvent gradient of 3%-97% B over 15 min was employed and the fractionated sample was directly subjected to ESI-Q/TOF MS analysis using a Waters Synapt mass spectrometer. The mass spectrometer was calibrated before samples were applied. Data was acquired in ESI positive mode with the capillary voltage set to 3.0-3.5 kV. The ionization source and desolvation gas were heated to 120 ℃ and 300 ℃, respectively. Cone gas was set to 0 L/h and desolvation gas was set to 600 L/h. The transfer collision energy was set to 4 V for both MS and MSMS analyses. The trap collision energy was set to 6 V for MS analysis. For MS/MS analysis, a trap collision energy ramp ranging from 20-40 V was applied on multiply charged parent ions to achieve fragmentation. Suitable trap collision energy was determined by choosing the spectra where both fragment peaks and parent peak could be observed. [Glu^1]-Fibrinopeptide B (Sigma) was directly infused as lock mass with lock spray

sampling if desired. The acquired spectra were processed using MaxEnt3 software and analyzed by Protein/Peptide Editor in BioLynx 4.1 (Waters).

Antimicrobial activity assay of CylL$_L$" and CylL$_S$"

CylL$_L$" and CylL$_S$" were obtained as described above by trypsin cleavage of CylM-modified CylL$_L$ E−1K and CylL$_S$ E−1K followed by HPLC purification. Peptides were dissolved in 2:1 acetonitrile:H_2O to achieve a concentration of 100 μM. The peptide solutions were diluted with phosphate buffered saline (PBS) to prepare a 10 μM solution. *L. lactis* HP cells were grown in GM17 media under anaerobic conditions at 25 ℃ for 16 h. Agar plates were prepared by combining 20 mL of melted GM17 agar (cooled to 42 °C for 5 min) with 200 μL of dense overnight cell culture. The seeded agar was poured into a sterile 100 mm round dish (VWR) and allowed to solidify at 25 ℃ for 30 min. Samples were directly spotted on the solidified agar. Plates were incubated at 30 ℃ for 16 h and the antimicrobial activity was determined by the presence or absence of zones of growth inhibition. A negative control was conducted using solutions with identical make-up but leaving out the peptides.

Hemolytic assay of CylL$_L$" and CylL$_S$"

A sample of 1 mL of defibrinated rabbit blood was diluted with 20 mL of PBS in a 50 mL conical tube on ice and mixed well by gently inverting the tube. The PBS-diluted blood sample was centrifuged at 1,000×*g* for 5 min at 4 ℃ and the supernatant containing lysed blood cells and released hemoglobin was discarded. The process was repeated 2 to 4 times until the supernatant was clear. The blood cells were then diluted with PBS to make a 5% solution, which was immediately used to test the hemolytic activity of the peptides. Rabbit blood was used within 10 days after receipt to minimize autolysis. CylL$_L$" and CylL$_S$" peptide solutions made as described above were diluted with PBS to prepare a 10 μM solution. To a 96-well plate, 100 μL of 5% washed rabbit blood sample was added to each well followed by the addition of the desired peptide samples or controls. In each sample well, CylL$_L$" was added and mixed well for 5 min before CylL$_S$" was added. PBS was used to adjust the final volume to 150 μL. The 96-well plate was kept in a 37 ^{o}C incubator to allow the lytic reaction to proceed. At each time point, 20 μL of each reaction mixture was taken out, diluted with 200 μL of fresh PBS and centrifuged at 1,000×*g* for 5

min. The supernatant (120 μL) was transferred to a new well and the absorbance was measured at 415 nm (wave length was optimized for rabbit hemoglobin). The absorbance of prepared blood sample at each time point was analyzed in triplicate and the maximum absorbance was determined by adding 50 μL of 0.1% Triton in PBS to 100 μL of 5% blood sample and using the same analysis procedure.

Purification of Halβ from *Bacillus halodurans* C-125

The isolation of Halβ from *Bacillus halodurans* C-125 was performed according to a previously reported procedure (*30*).

GC-MS analysis

The synthesis of Lan and MeLan standards and the preparation of samples for GC-MS analysis were carried out following a reported procedure published elsewhere (*20, 21*). The NMR spectra of the synthetic standards matched the literature values. The modified precursor peptides with their leader sequences attached were hydrolyzed and the resulting solutions were dried and directly used for derivatization. The derivatized samples were analyzed by GC-MS using an Agilent 7890 gas chromatograph equipped with a Varian CP-Chirasil-L-Val fused silica column (25 m x 0.25 mm x 0.15 μm). Samples were dissolved in methanol and introduced to the instrument via a splitless/split (1:20) injection at a flow rate of 1.7 or 2.0 mL/min helium gas. The temperature method used was 160 ^{o}C for 5 min, 160 ^{o}C to 180 ^{o}C at 3 ^{o}C/min, and 180 ^{o}C for 6 min. The mass spectrometer was operated in simultaneous scan/selected-ion monitoring (SIM) mode, monitoring at the characteristic fragment masses of 365 Da for Lan and 379 Da for MeLan residues.

2.4 References

1. Gilmore, M. S., Segarra, R. A., Booth, M. C., Bogie, C. P., Hall, L. R., and Clewell, D. B. (1994) Genetic structure of the *Enterococcus faecalis* plasmid pAD1-encoded cytolytic toxin system and its relationship to lantibiotic determinants., *J. Bacteriol. 176*, 7335-7344.
2. Cox, C. R., Coburn, P. S., and Gilmore, M. S. (2005) Enterococcal cytolysin: A novel two component peptide system that serves as a bacterial defense against eukaryotic and prokaryotic cells, *Curr. Protein Pept. Sci. 6*, 77-84.
3. Lawton, E. M., Ross, R. P., Hill, C., and Cotter, P. D. (2007) Two-peptide lantibiotics: a medical perspective, *Mini Rev. Med. Chem. 7*, 1236-1247.
4. Bierbaum, G., and Sahl, H. G. (2009) Lantibiotics: mode of action, biosynthesis and bioengineering, *Curr. Pharm. Biotechnol. 10*, 2-18.
5. Ike, Y., and Clewell, D. B. (1984) Genetic analysis of the pAD1 pheromone response in *Streptococcus faecalis*, using transposon Tn917 as an insertional mutagen, *J. Bacteriol. 158*, 777-783.
6. Chow, J. W., Thal, L. A., Perri, M. B., Vazquez, J. A., Donabedian, S. M., Clewell, D. B., and Zervos, M. J. (1993) Plasmid-associated hemolysin and aggregation substance production contribute to virulence in experimental enterococcal endocarditis, *Antimicrob. Agents Chemother. 37*, 2474-2477.
7. Huycke, M. M., Spiegel, C. A., and Gilmore, M. S. (1991) Bacteremia caused by hemolytic, high-level gentamicin-resistant *Enterococcus faecalis*, *Antimicrob. Agents Chemother. 35*, 1626-1634.
8. Coburn, P. S., and Gilmore, M. S. (2003) The *Enterococcus faecalis* cytolysin: a novel toxin active against eukaryotic and prokaryotic cells, *Cell. Microbiol. 5*, 661-669.
9. Haas, W., Shepard, B. D., and Gilmore, M. S. (2002) Two-component regulator of Enterococcus faecalis cytolysin responds to quorum-sensing autoinduction, *Nature 415*, 84-87.
10. Coburn, P. S., Pillar, C. M., Jett, B. D., Haas, W., and Gilmore, M. S. (2004) *Enterococcus faecalis* senses target cells and in response expresses cytolysin, *Science 306*, 2270-2272.
11. Knerr, P. J., and van der Donk, W. A. (2012) Discovery, biosynthesis, and engineering of lantipeptides, *Annu. Rev. Biochem. 81*, 479-505.
12. Haas, W., and Gilmore, M. S. (1999) Molecular nature of a novel bacterial toxin: the cytolysin of Enterococcus faecalis, *Med. Microbiol. Immunol. 187*, 183-190.
13. Gilmore, M. S., Segarra, R. A., and Booth, M. C. (1990) An HlyB-type function is required for expression of the Enterococcus faecalis hemolysin/bacteriocin, *Infect. Immun. 58*, 3914-3923.
14. Segarra, R. A., Booth, M. C., Morales, D. A., Huycke, M. M., and Gilmore, M. S. (1991) Molecular characterization of the Enterococcus faecalis cytolysin activator, *Infect. Immun.*

59, 1239-1246.

15. Booth, M. C., Bogie, C. P., Sahl, H. G., Siezen, R. J., Hatter, K. L., and Gilmore, M. S. (1996) Structural analysis and proteolytic activation of *Enterococcus faecalis* cytolysin, a novel lantibiotic, *Mol. Microbiol. 21*, 1175-1184.
16. Shi, Y., Yang, X., Garg, N., and van der Donk, W. A. (2011) Production of lantipeptides in *Escherichia coli*, *J. Am. Chem. Soc. 133*, 2338-2341.
17. Garg, N., Tang, W., Goto, Y., Nair, S. K., and van der Donk, W. A. (2012) Lantibiotics from *Geobacillus thermodenitrificans*, *Proc. Nat. Acad. Sci. U. S. A. 109*, 5241-5246.
18. McClerren, A. L., Cooper, L. E., Quan, C., Thomas, P. M., Kelleher, N. L., and van der Donk, W. A. (2006) Discovery and in vitro biosynthesis of haloduracin, a two-component lantibiotic, *Proc. Nat. Acad. Sci. U. S. A. 103*, 17243-17248.
19. Chatterjee, C., Paul, M., Xie, L. L., and van der Donk, W. A. (2005) Biosynthesis and mode of action of lantibiotics, *Chem. Rev. 105*, 633-683.
20. Liu, W., Chan, A. S. H., Liu, H. Q., Cochrane, S. A., and Vederas, J. C. (2011) Solid supported chemical syntheses of both components of the lantibiotic lacticin 3147, *J. Am. Chem. Soc. 133*, 14216-14219.
21. Tang, W., and van der Donk, W. A. (2012) Structural characterization of four prochlorosins: a novel class of lantipeptides produced by planktonic marine cyanobacteria, *Biochemistry 51*, 4271-4279.
22. Kido, Y., Hamakado, T., Yoshida, T., Anno, M., Motoki, Y., Wakamiya, T., and Shiba, T. (1983) Isolation and characterization of ancovenin, a new inhibitor of angiotensin I converting enzyme, produced by actinomycetes, *J. Antibiot. 36*, 1295-1299.
23. Hyink, O., Balakrishnan, M., and Tagg, J. R. (2005) *Streptococcus rattus* strain BHT produces both a class I two-component lantibiotic and a class II bacteriocin, *FEMS Microbiol. Lett. 252*, 235-241.
24. Yonezawa, H., and Kuramitsu, H. K. (2005) Genetic analysis of a unique bacteriocin, Smb, produced by *Streptococcus mutans* GS5, *Antimicrob. Agents Chemother. 49*, 541-548.
25. Ike, Y., Clewell, D. B., Segarra, R. A., and Gilmore, M. S. (1990) Genetic analysis of the pAD1 hemolysin/bacteriocin determinant in *Enterococcus faecalis*: Tn917 insertional mutagenesis and cloning, *J. Bacteriol. 172*, 155-163.
26. Cooper, L. E., McClerren, A. L., Chary, A., and van der Donk, W. A. (2008) Structure-Activity Relationship Studies of the Two-Component Lantibiotic Haloduracin, *Chem. Biol. 15*, 1035-1045.
27. Ross, A. C., Liu, H., Pattabiraman, V. R., and Vederas, J. C. (2010) Synthesis of the lantibiotic lactocin S using peptide cyclizations on solid phase, *J. Am. Chem. Soc. 132*, 462-463.
28. Ross, A. C., and Vederas, J. C. (2011) Fundamental functionality: recent developments in understanding the structure-activity relationships of lantibiotic peptides, *J. Antibiot. (Tokyo)*

64, 27-34.

29. Li, B., Cooper, L. E., and van der Donk, W. A. (2009) *In vitro* studies of lantibiotic biosynthesis *Methods Enzymol. 458*, 533-558.
30. Oman, T. J., and van der Donk, W. A. (2009) Insights into the Mode of Action of the Two-Peptide Lantibiotic Haloduracin, *ACS Chem. Biol. 4*, 865-874.

Chapter 3. Substrate Control in Stereoselective Lanthionine Biosynthesis

3.1 Introduction

The stereochemical outcome of chemical reactions between chiral molecules in which one or more new stereocenters are generated is often governed by the stereochemistry of the substrate or reagent, resulting in substrate or reagent controlled processes (*1*). For enzymatic reactions, the well-organized chiral environment of the enzyme active site typically results in reagent control for physiological processes. In this chapter, I describe a rare case where the substrate controls the stereochemical outcome of an enzyme-catalyzed Michael-type addition during the biosynthesis of lanthipeptides (for an introduction about lanthipeptides, see section 1.2 of chapter 1). The biosynthetic route of class II lanthipeptides is illustrated in **Figure 3.1a** using the enterococcal cytolysin as an example.

Lanthipeptides have been extensively studied over the past 30 years, with structural information available for a subset of compounds (*2*). Until recently, it was generally assumed that the cysteine addition always gave (2*S*, 6*R*)-Lan (hereafter called DL-Lan) and (2*S*, 3*S*, 6*R*)-MeLan (hereafter DL-MeLan) (**Figure 3.1b**). This assumption was recently challenged by the discovery of (2*R*, 3*R*, 6*R*)-MeLan (LL-MeLan, **Figure 3.1b**) in two class II lantibiotics - the cytolysin from *Enterococcus faecalis* and haloduracin produced by *Bacillus halodurans* (see chapter 2) (*3*). The enterococcal cytolysin consists of two peptides called cytolysin L ($CylL_L$'', **Figure 3.1c**) and cytolysin S ($CylL_S$'', **Figure 3.1a**), which are generated from the precursor peptides $CylL_L$ and $CylL_S$, respectively, by a single lanthionine synthetase CylM. The production of cytolysin enhances virulence in infection models of *E. faecalis* and is associated with acute patient mortality (*4-7*). Haloduracin is an antimicrobial substance that also consists of two peptides (**Figure 3.1c**) (*8*). Based on sequence homology of the rings containing the LL-MeLan residues in cytolysin and haloduracin, it has been suggested that a Dhb-Dhb-Xxx-Xxx-Cys motif (Xxx represents amino acids other than Dha, Dhb and Cys) (**Figure 3.1**) induces the formation of the unusual MeLan

---

Reproduced in part with the permission from:
Tang, W., Jiménez-Osés, G., Houk, K. N., and van der Donk, W. A. (2015) Substrate control in stereoselective lanthionine biosynthesis, *Nat. Chem. 7*, 57-64.

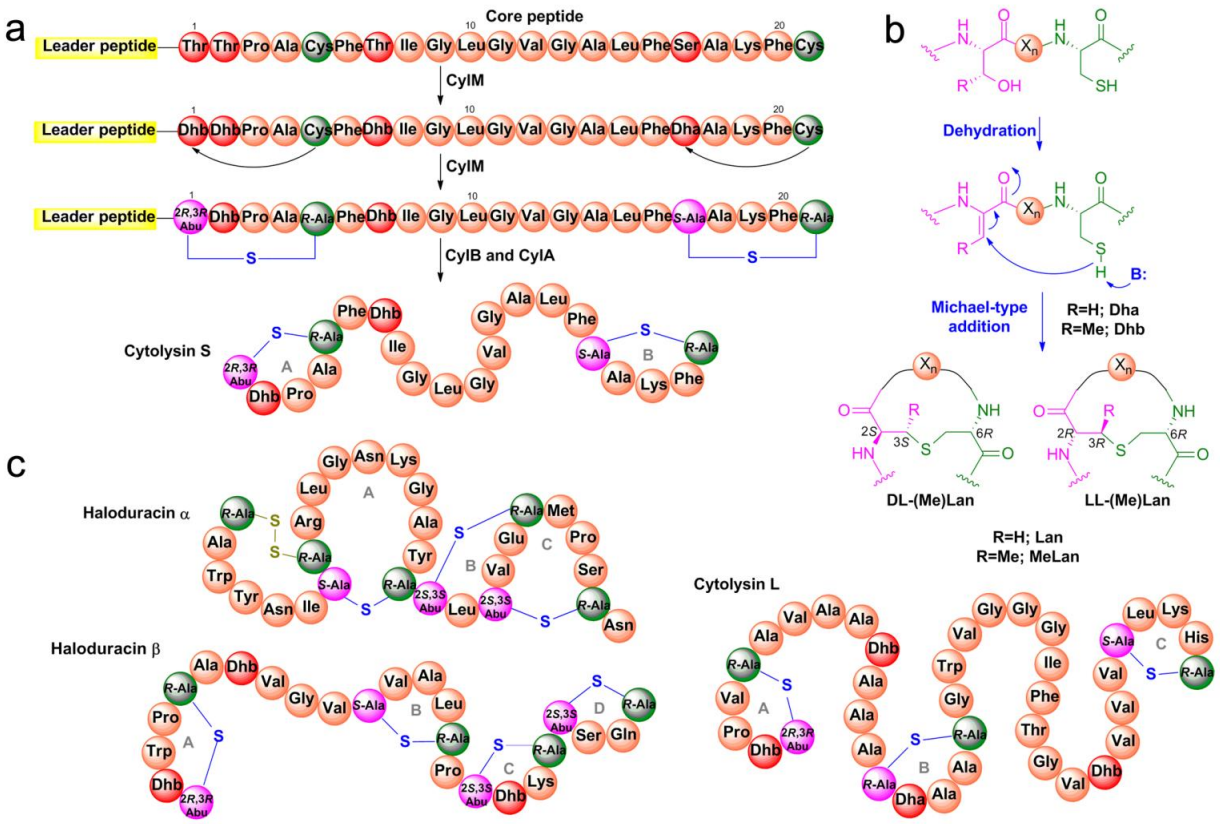

Figure 3.1 Biosynthesis of class II lanthipeptides. (a) Generic pathway of class II lanthipeptide biosynthesis with cytolysin S as an example. The $CylL_S$ peptide is synthesized by the ribosome as a linear precursor peptide with an N-terminal leader peptide and a C-terminal core peptide. The dehydration of Ser and Thr residues to Dha and Dhb residues as well as the Michael-type addition by Cys residues onto these α,β-unsaturated amino acids is catalyzed by the bi-functional lanthionine synthetase CylM. The leader peptide is then removed by the proteases CylB and CylA, affording mature cytolysin S. **(b)** Dehydration and cyclization reactions catalyzed by LanM enzymes that result in (Me)Lan residues with the canonical DL (i.e. D configuration at the α carbon originating from Ser or Thr and L configuration at the α carbon of the former Cys) and unusual LL (L configuration for both α carbons) stereochemistries. **(c)** The structures of haloduracin (Halα and Halβ) and cytolysin L. Abu, α-aminobutyric acid.

stereochemistry (*3*). In this model, the LL stereochemistry is formed by *anti*-addition of cysteine to the *Re* face of the alkene because of a conformational preference of the substrate that is unique to this motif (*3*).

These findings have raised a number of intriguing questions. First, given that cytolysin and haloduracin contain both LL- and DL-(Me)Lan, how can one enzyme catalyze similar conjugate additions with different stereochemistries in a single polypeptide substrate? Second, does the substrate sequence indeed govern the diastereoselectivity? And third, is it possible to manipulate

the stereochemical outcome by mutating the peptide substrate? In this chapter, I show that the formation of LL-MeLan from the Dhb-Dhb-Xxx-Xxx-Cys sequence is not specific to the enzymes involved in cytolysin or haloduracin biosynthesis. Furthermore, I demonstrate that a Dhb-to-Ala mutation at the second position of the motif results in MeLan formation with the canonical DL stereochemistry, supporting the substrate control hypothesis. Quantum mechanical simulations of dehydrated peptides (performed by Dr. Gonzalo Jiménez-Osés and Prof. K. N. Houk, Department of Chemistry and Biochemistry, University of California, Los Angeles) confirm an intrinsic preference for *Re* face or *Si* face addition depending on the sequence of the substrate peptide.

3.2 Results

3.2.1 Generality of LL-MeLan formation from Dhb-Dhb-Xxx-Xxx-Cys

To test the hypothesis that the substrate peptide sequence containing the Dhb-Dhb-Xxx-Xxx-Cys motif determines the formation of LL stereochemistry independent of the identity of the lanthionine synthetase, a series of chimeric peptides were generated with the $CylL_S$ core peptide connected to leader peptides from different class II lanthipeptides. These leader peptides were expected to be recognized by their cognate lanthionine synthetases that carry out the dehydration and cyclization reactions in the core peptide. The gene *halA2-cylL$_S$* that encodes a chimeric peptide consisting of the leader peptide of the haloduracin α precursor peptide (HalA2) fused to the $CylL_S$ core peptide (**Figure 3.2a**) was co-expressed in *E. coli* with *halM2* encoding the Halβ synthetase HalM2 (*9*). Modified HalA2-$CylL_S$ was purified and analysis by matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectrometry (MS) showed the desired four dehydrations (**Figure 3.2b**). The modified chimeric peptide was hydrolyzed in 6 M HCl to individual amino acid residues, which were derivatized and analyzed by gas chromatography (GC) MS using a chiral GC column to reveal the stereochemistry of the (Me)Lan residues as described in chapter 2 (*3, 10*). Signals corresponding to LL-MeLan and DL-Lan were observed as the predominant products by comparison with synthetic standards derivatized in the same way (**Figure 3.3** and **Table 3.1**). Hence, HalM2 catalyzes the cyclization of the $CylL_S$ core peptide with the same stereochemical selectivity as CylM.

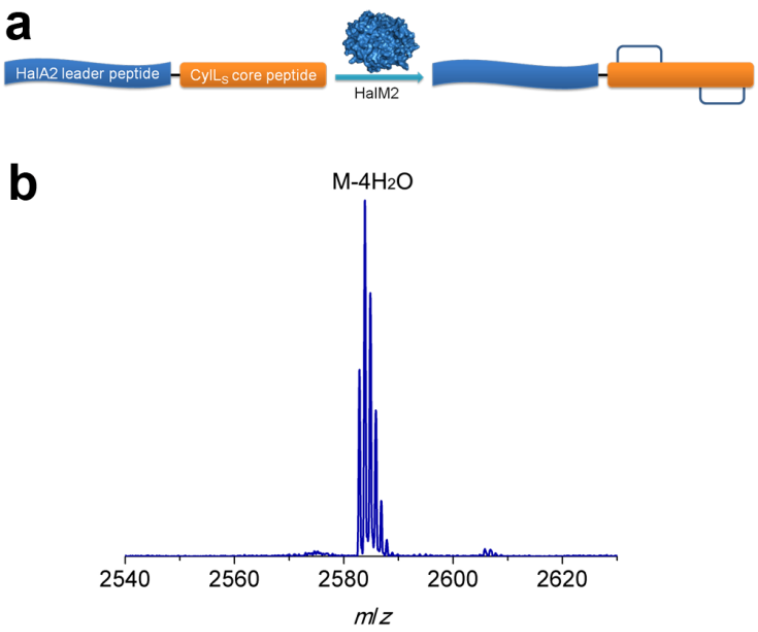

Figure 3.2 Chimeric peptide HalA2-CylL$_S$ modified by HalM2 in *E. coli*. (a) HalA2-CylL$_S$ chimeric peptide that was modified by the Halβ synthetase HalM2 in *E. coli*. A generic protein structure is shown for the enzyme. **(b)** MALDI/TOF mass spectrum of HalM2-modified HalA2-CylL$_S$ core peptide. The core peptide contains five additional amino acids at its N terminus originating from the leader peptide as a result of AspN cleavage (calculated M−4 H_2O: 2581.3, monoisotopic; observed: 2582.9, monoisotopic, M−4 H_2O+H^+).

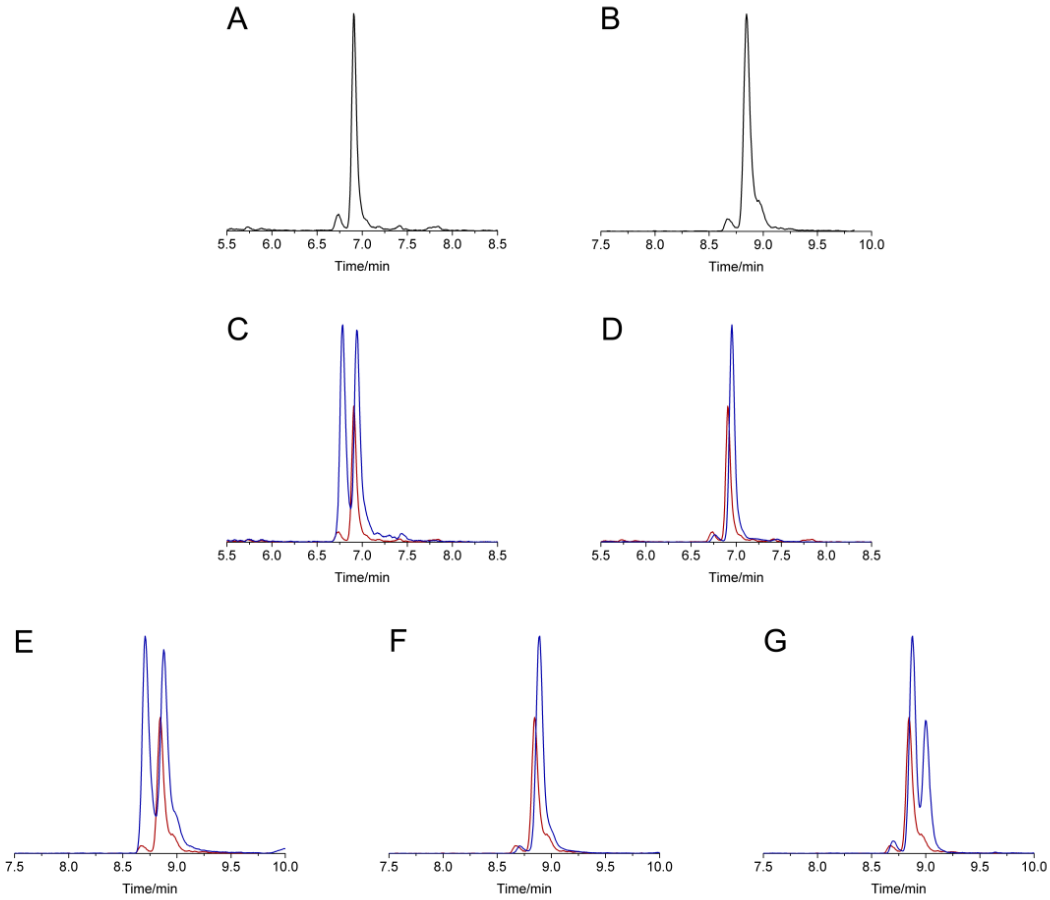

Figure 3.3 GC-MS traces for derivatized Lan/MeLan residues obtained from hydrolysis of HalA2-CylL$_S$ modified by HalM2 in *E. coli* and co-injections with derivatized Lan/MeLan standards (selected ion monitoring, SIM, at 365 Da for Lan and 379 Da for MeLan). **(a)** Hydrolyzed and derivatized MeLan residues from modified HalA2-CylL$_S$. **(b)** Hydrolyzed and derivatized Lan residues from modified HalA2-CylL$_S$. **(c)** Hydrolyzed and derivatized MeLan residues from modified HalA2-CylL$_S$ (red line) and co-injected with derivatized DL-MeLan standard (blue line). **(d)** Hydrolyzed and derivatized MeLan residues from modified HalA2-CylL$_S$ (red line) and co-injected with derivatized LL-MeLan standard (blue line). **(e)** Hydrolyzed and derivatized Lan residues from modified HalA2-CylL$_S$ (red line) and co-injected with derivatized DD-Lan standard (blue line). **(f)** Hydrolyzed and derivatized Lan residues from modified HalA2-CylL$_S$ (red line) and co-injected with derivatized DL-Lan standard (blue line). **(g)** Hydrolyzed and derivatized Lan residues from modified HalA2-CylL$_S$ (red line) and co-injected with derivatized LL-Lan standard (blue line). Traces of derivatized Lan and MeLan residues from modified HalA2-CylL$_S$ (red lines) used for overlay were adjusted to 70% intensity for clarity. A small amount of epimerization of the (Me)Lan occurs during acid hydrolysis of the peptides as has been reported previously (*11*).

This result was perhaps not too surprising since HalM2 converts a Dhb-Dhb-Trp-Pro-Cys sequence into an LL-MeLan cross-link in its native substrate HalA2 (**Figure 3.1c**) (*3*). Therefore, I turned to LanM enzymes that only form DL-(Me)Lan in their natural substrates. The lanthionine synthetase LtnM2 involved in the biosynthesis of the lantibiotic lacticin 3147 A2 (*12*) was evaluated next. An LtnA2-CylL$_S$ peptide consisting of the leader peptide from the natural LtnM2 substrate (LtnA2) and the core peptide from CylL$_S$ was co-expressed with LtnM2 in *E. coli*. MALDI-TOF MS analysis of the peptide product illustrated the desired four dehydrations (**Figure 3.4**). Hydrolysis and derivatization of the modified LtnA2-CylL$_S$ peptide once again resulted in almost exclusive LL-MeLan and DL-Lan signals in GC-MS experiments (**Figure 3.5** and **Table 3.1**), demonstrating that LtnM2 also preferentially forms the LL stereochemistry when guided by the substrate sequence.

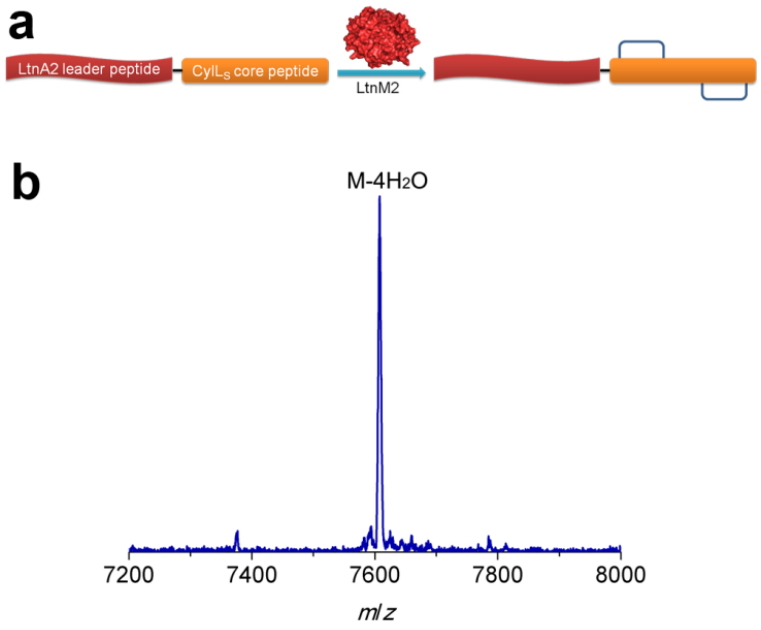

Figure 3.4 Chimeric peptide LtnA2-CylL$_S$ modified by LtnM2 in *E. coli*. (a) LtnA2-CylL$_S$ chimera that was modified by the Ltnβ synthetase LtnM2 in *E. coli*. A generic protein structure is shown for the enzyme. **(b)** MALDI/TOF mass spectrum of LtnM2-modified LtnA2-CylL$_S$ peptide (calculated M−4 H_2O: 7606; observed: 7607, M−4 H_2O+H^+).

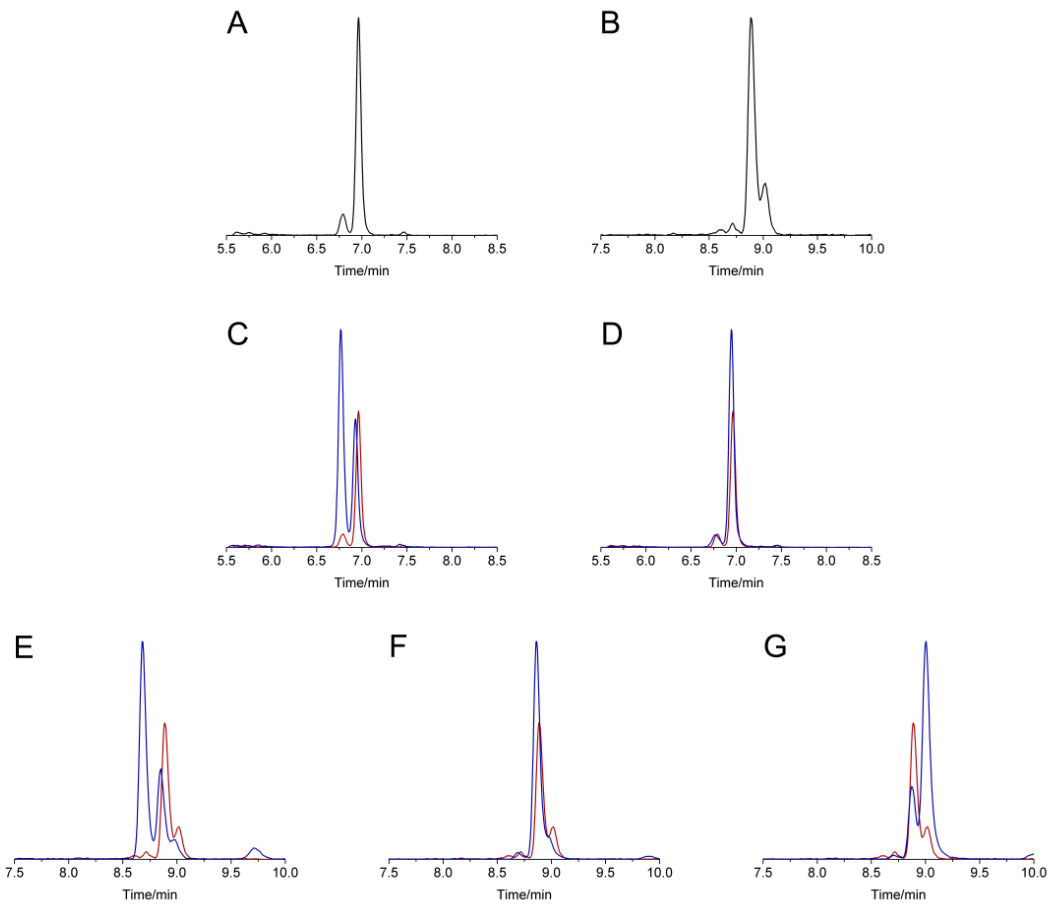

Figure 3.5 GC-MS traces for derivatized Lan/MeLan residues obtained from hydrolysis of LtnA2-CylL$_S$ modified by LtnM2 in *E. coli* and co-injections with derivatized Lan/MeLan standards (selected ion monitoring, SIM, at 365 Da for Lan and 379 Da for MeLan). **(a)** Hydrolyzed and derivatized MeLan residues from modified LtnA2-CylL$_S$. **(b)** Hydrolyzed and derivatized Lan residues from modified LtnA2-CylL$_S$. **(c)** Hydrolyzed and derivatized MeLan residues from modified LtnA2-CylL$_S$ (red line) and co-injected with derivatized DL-MeLan standard (blue line). **(d)** Hydrolyzed and derivatized MeLan residues from modified LtnA2-CylL$_S$ (red line) and co-injected with derivatized LL-MeLan standard (blue line). **(e)** Hydrolyzed and derivatized Lan residues from modified LtnA2-CylL$_S$ (red line) and co-injected with derivatized DD-Lan standard (blue line). **(f)** Hydrolyzed and derivatized Lan residues from modified LtnA2-CylL$_S$ (red line) and co-injected with derivatized DL-Lan standard (blue line). **(g)** Hydrolyzed and derivatized Lan residues from modified LtnA2-CylL$_S$ (red line) and co-injected with derivatized LL-Lan standard (blue line). Traces of derivatized Lan and MeLan residues from modified LtnA2-CylL$_S$ (red lines) used for overlay were adjusted to 70% intensity for clarity. A small amount of epimerization of the (Me)Lan occurs during acid hydrolysis of the peptides as has been reported previously (*11*).

I then chose ProcM as the next LanM target. ProcM is a lanthionine synthetase with 30 natural substrates (ProcAs) that exhibits high substrate tolerance (**Figure 3.6**, also see chapter 8) (*13*). ProcM can generate many different ring topologies, and based on the currently structurally characterized ProcA products, they all have the DL-(Me)Lan stereochemistry (*14*). It has been suggested that it is the sequence of the ProcA peptides, rather than the ProcM enzyme, that primarily determines the cyclization patterns (*15*). As such, ProcM would be the perfect candidate to test the hypothesis that the unusual LL-MeLan stereochemistry produced from the Dhb-Dhb-Xxx-Xxx-Cys motif is determined by the substrate. The ProcA3.2 leader peptide was chosen for a chimeric ProcA3.2-CylL_S peptide and the latter was co-expressed with ProcM in *E. coli* (**Figure 3.7a**). The modified ProcA3.2-CylL_S peptide afforded several products that differed in the extent of dehydration with 2- and 4-fold dehydrated peptides as the major products (**Figure 3.7b**). The modified peptides were hydrolyzed, and the resulting amino acids were derivatized and subjected to GC-MS analysis. In this case, an LL-MeLan peak with a relatively low intensity as well as a DL-Lan signal was observed (**Figure 3.8** and **Table 3.1**). The low signal intensity of the derivatized MeLan compared with Lan could be the result of the observed incomplete dehydration of the non-cognate core peptide or because of partial cyclization by ProcM. Regardless, the MeLan that was generated clearly was formed with high selectivity for the LL stereochemistry.

Collectively, these results demonstrate that class II LanM enzymes have a general preference for converting sequences containing the Dhb-Dhb-Xxx-Xxx-Cys motif into LL-MeLan (**Figure 3.9** and **Table 3.1**).

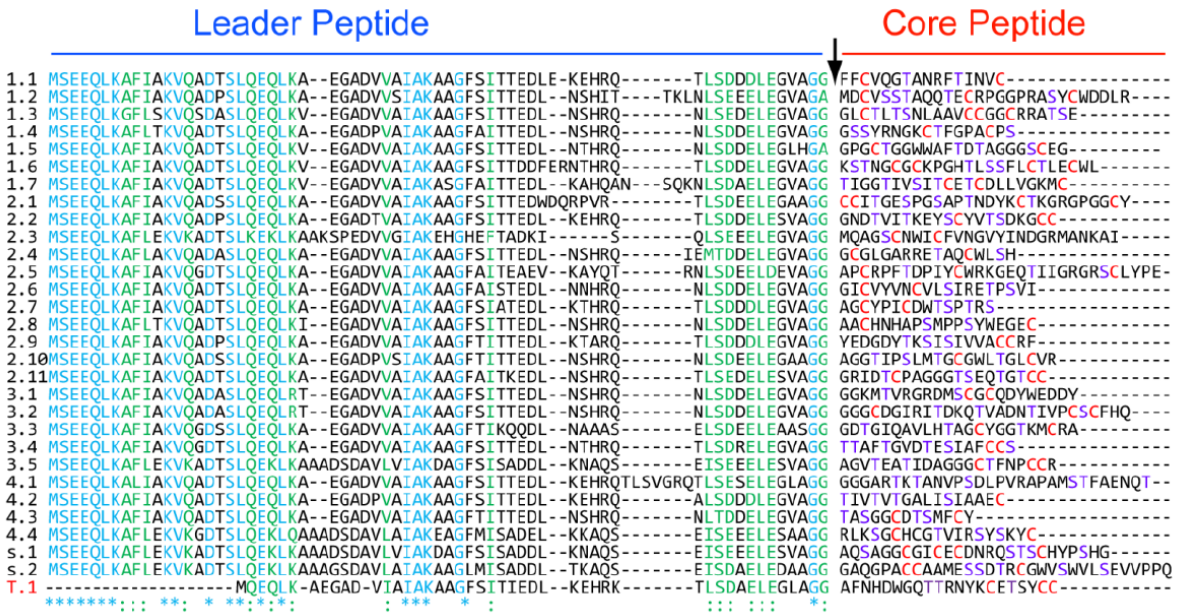

Figure 3.6 Sequences of 30 ProcAs encoded by the genome of Prochlorococcus MIT 9313 (*16*).

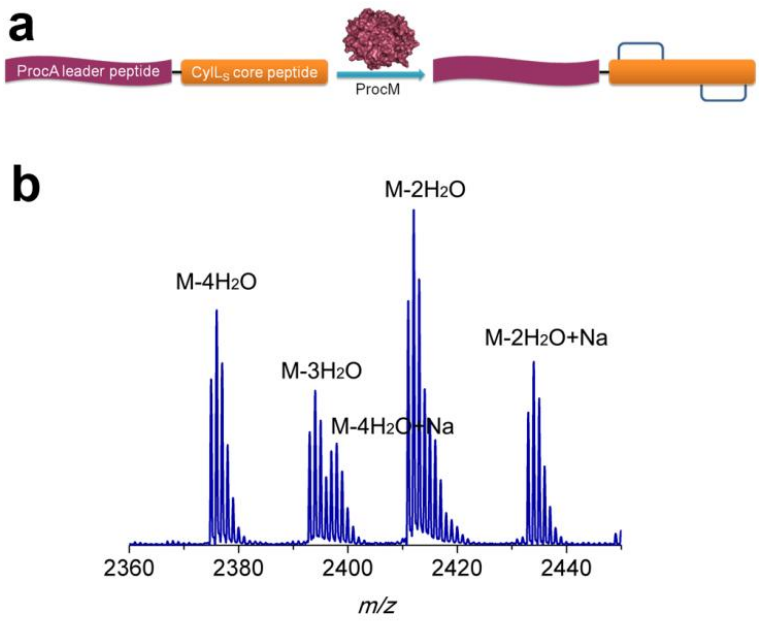

Figure 3.7 Chimeric peptide ProcA3.2-CylL$_S$ modified by ProcM in *E. coli*. (a) ProcA3.2-CylL$_S$ that was modified by the ProcA synthetase ProcM in *E. coli*. A generic protein structure is shown for the enzyme. **(b)** MALDI/TOF mass spectrum of ProcM-modified ProcA3.2-CylL$_S$ core peptide. The core peptide contains five additional amino acids at its N terminus originating from the leader peptide after endoproteinase GluC cleavage (calculated M−4 H_2O: 2372.2, monoisotopic; observed: 2374.1, monoisotopic, M−4 H_2O+H^+). The 0.9 Da mass difference is presumably introduced by application of external calibration instead of internal calibration of peptide standards.

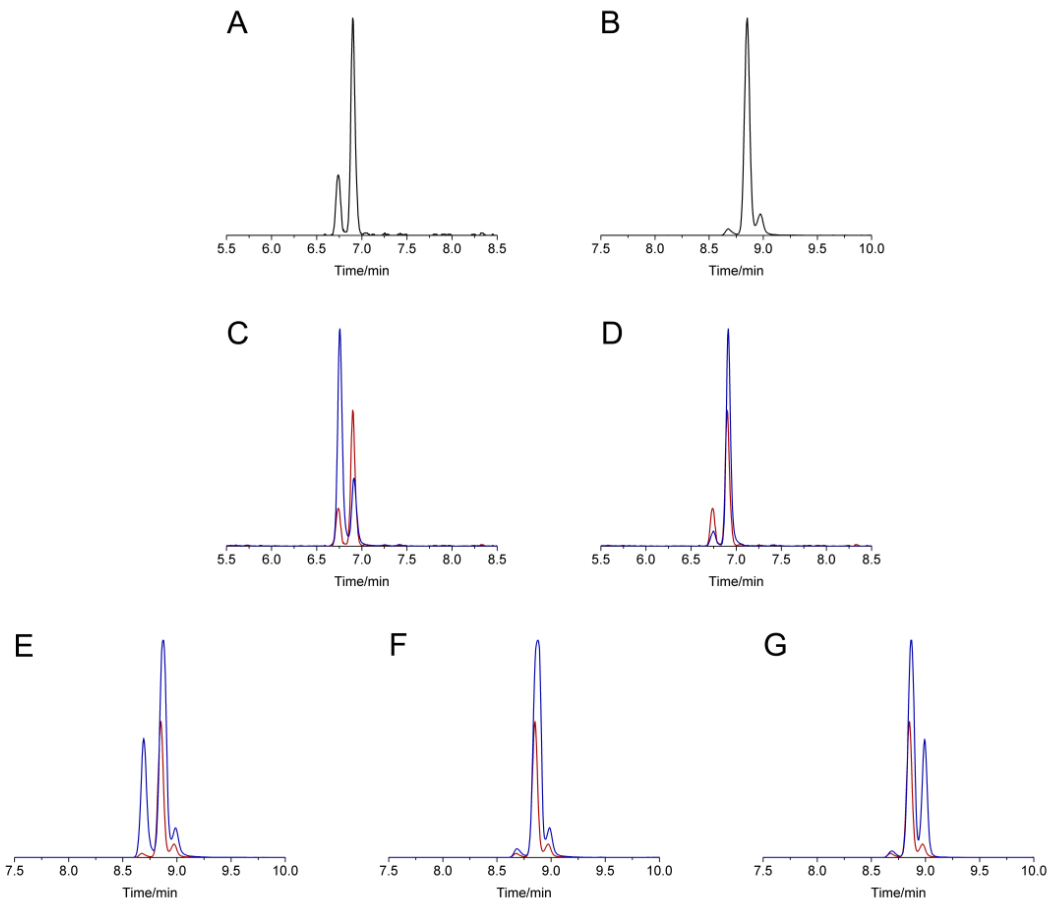

Figure 3.8 GC-MS traces for derivatized Lan/MeLan residues obtained from hydrolysis of ProcA3.2-CylL$_S$ modified by ProcM in *E. coli* and co-injections with derivatized Lan/MeLan standards (selected ion monitoring, SIM, at 365 Da for Lan and 379 Da for MeLan). **(a)** Hydrolyzed and derivatized MeLan residues from modified ProcA3.2-CylL$_S$. **(b)** Hydrolyzed and derivatized Lan residues from modified ProcA3.2-CylL$_S$. **(c)** Hydrolyzed and derivatized MeLan residues from modified ProcA3.2-CylL$_S$ (red line) and co-injected with derivatized DL-MeLan standard (blue line). **(d)** Hydrolyzed and derivatized MeLan residues from modified ProcA3.2-CylL$_S$ (red line) and co-injected with derivatized LL-MeLan standard (blue line). **(e)** Hydrolyzed and derivatized Lan residues from modified ProcA3.2-CylL$_S$ (red line) and co-injected with derivatized DD-Lan standard (blue line). **(f)** Hydrolyzed and derivatized Lan residues from modified ProcA3.2-CylL$_S$ (red line) and co-injected with derivatized DL-Lan standard (blue line). **(g)** Hydrolyzed and derivatized Lan residues from modified ProcA3.2-CylL$_S$ (red line) and co-injected with derivatized LL-Lan standard (blue line). Traces of derivatized Lan and MeLan residues from modified ProcA3.2-CylL$_S$ (red lines) used for overlay were adjusted to 70% intensity for clarity. A small amount of epimerization of the (Me)Lan occurs during acid hydrolysis of the peptides as has been reported previously (*11*).

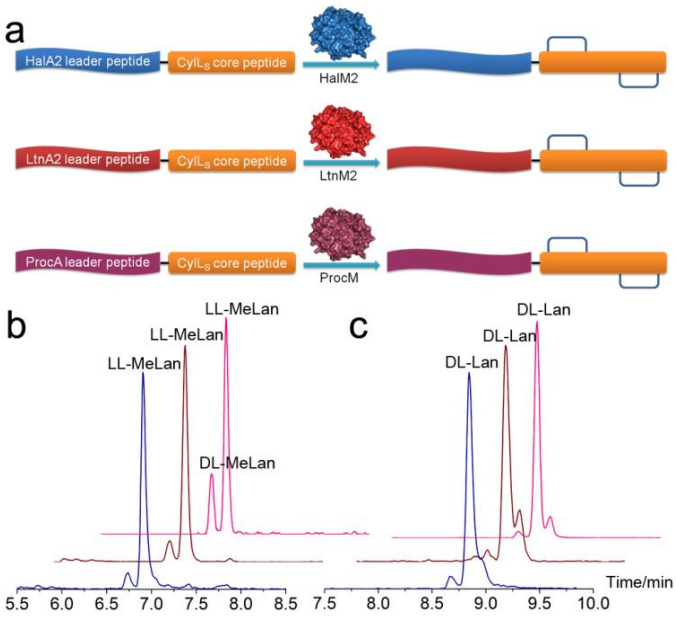

Figure 3.9 GC-MS analysis of *in vivo* modified HalA2-CylL$_S$, LtnA2-CylL$_S$ and ProcA3.2-CylL$_S$ chimeric peptides. (a) HalA2-CylL$_S$, LtnA2-CylL$_S$ and ProcA3.2-CylL$_S$ chimeras that were modified in *E. coli* by the lanthionine synthetases HalM2, LtnM2, and ProcM, respectively. A generic protein structure is shown for the three enzymes. GC-MS traces are shown for hydrolyzed and derivatized MeLan residues **(b)** and Lan residues **(c)** from modified HalA2-CylL$_S$ (blue), LtnA2-CylL$_S$ (red), and ProcA3.2-CylL$_S$ (magenta). All intensities were normalized with respect to the predominant product.

3.2.2 Mutation of the Dhb-Dhb-Xxx-Xxx-Cys motif

To probe whether the two consecutive dehydro amino acids in the motif impart formation of the LL stereochemistry, Thr2 in HalA2 was mutated to Ala. The HalA2-T2A peptide was co-expressed with its synthetase HalM2 in *E. coli*, resulting in the desired elimination of six water molecules. After leader peptide removal via an engineered Factor Xa cleavage site (*9*), the modified HalA2-T2A core peptide was subjected to tandem MS (MS/MS) analysis, which confirmed the formation of the expected ring pattern (**Figure 3.10a**). The modified peptide was then hydrolyzed, and the resulting amino acids were derivatized and analyzed by GC-MS. A DL-MeLan signal was almost exclusively observed, compared with a 2:1 ratio of DL:LL MeLan signals observed for the wild-type HalA2 peptide after HalM2 modification (the LL-MeLan residue originating from the A ring and two DL-MeLan residues originating from the C and D rings) (**Figures 3.10b** and **3.11**; **Table 3.1**) (*3*); as expected the DL-configuration of the Lan residue (ring B) remained unchanged in the mutant (**Figure 3.11**). To rule out the possibility that the A ring of the HalA2-T2A mutant was not properly cyclized, which would mean that the observed DL-MeLan signal originated solely from the C and D rings, I used trypsin to remove the

leader peptide and analyzed the core peptide mass with liquid chromatography MS (LC-MS). If the A ring of the HalA2-T2A mutant was not properly cyclized, the core peptide would start with a Dhb, which would hydrolyze to a ketone after the removal of the leader peptide, giving a mass increase of 1 Da (*8*). Trypsin was chosen over Factor Xa because trypsin exhibits minimal bias in terms of digestion efficiency when the cleavage site is located next to a dehydrated amino acid (*17*). The observed mass for the HalA2-T2A core peptide by electrospray ionization (ESI) MS was 2318.0476 Da (**Figure 3.10c**), in agreement with the calculated mass of the cyclized peptide (2318.0469 Da). No peak with a mass corresponding to the non-cyclized peptide (calculated 2319.0309 Da) was observed in the LC trace. The mass spectrometer was calibrated before samples were applied and an internal standard sample was directly infused when analyzing the peptide samples to correct any possible mass drift. Collectively, these results show that the A ring of the HalA2-T2A mutant was successfully cyclized and had the DL stereochemistry, strongly suggesting that the 2nd Dhb in the Dhb-Dhb-Trp-Pro-Cys sequence is critical for the formation of LL-MeLan by HalM2. By mutation of the 2nd Dhb to Ala, the stereochemistry of the MeLan residue in the A ring of HalA2 could be successfully changed from LL to DL.

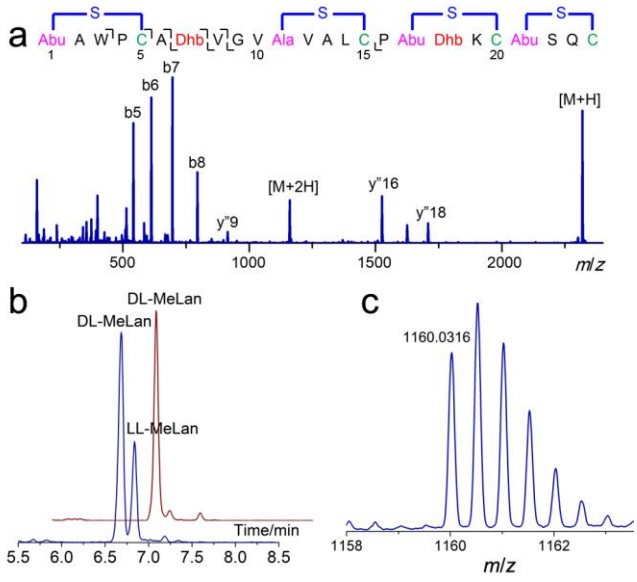

Figure 3.10 Structural characterization of HalM2-modified HalA2-T2A peptide. (a) MS/MS analysis of the HalA2-T2A core peptide modified by HalM2 in *E. coli*. Fragmentation is suppressed in the sequences spanning residues 1-5, 11-15, 17-20 and 21-24 as a result of ring formation. The b3 ion is attributed to in-ring fragmentation, which is rare but has also been observed for cytolysin, nisin and geobacillin (*3*, *9*, *18*). **(b)** GC-MS traces for hydrolyzed and derivatized MeLan residues from HalM2-modified HalA2 peptide (blue) and HalM2-modified HalA2-T2A peptide (red). **(c)** Electrospray ionization (ESI) quadrupole TOF mass spectrum for modified HalA2-T2A core peptide. The $[M+2H]^{2+}$ ion was observed with an *m*/*z* value of 1160.0316 (calculated *m*/*z* of cyclized HalA2-T2A core peptide with 2 protons: 1160.0313).

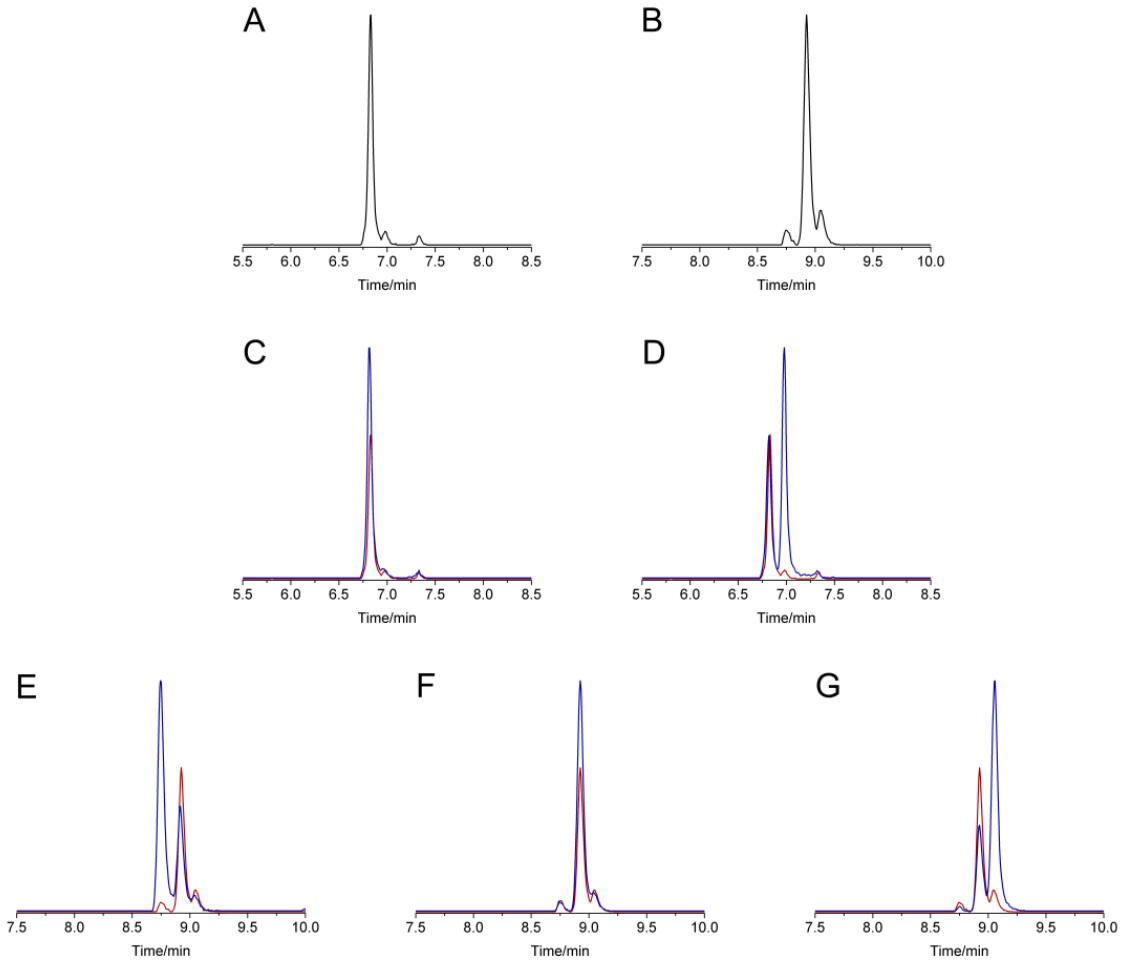

Figure 3.11 GC-MS traces for derivatized Lan/MeLan residues obtained from hydrolysis of HalA2-T2A modified by HalM2 in *E. coli* and co-injections with derivatized Lan/MeLan standards (selected ion monitoring, SIM, at 365 Da for Lan and 379 Da for MeLan). **(a)** Hydrolyzed and derivatized MeLan residues from modified HalA2-T2A. **(b)** Hydrolyzed and derivatized Lan residues from modified HalA2-T2A. **(c)** Hydrolyzed and derivatized MeLan residues from modified HalA2-T2A (red line) and co-injected with derivatized DL-MeLan standard (blue line). **(d)** Hydrolyzed and derivatized MeLan residues from modified HalA2-T2A (red line) and co-injected with derivatized LL-MeLan standard (blue line). **(e)** Hydrolyzed and derivatized Lan residues from modified HalA2-T2A (red line) and co-injected with derivatized DD-Lan standard (blue line). **(f)** Hydrolyzed and derivatized Lan residues from modified HalA2-T2A (red line) and co-injected with derivatized DL-Lan standard (blue line). **(g)** Hydrolyzed and derivatized Lan residues from modified HalA2-T2A (red line) and co-injected with derivatized LL-Lan standard (blue line). Traces of derivatized Lan and MeLan residues from modified HalA2-T2A (red lines) used for overlay were adjusted to 70% intensity for clarity. A small amount of epimerization of the (Me)Lan occurs during acid hydrolysis of the peptides as has been reported previously (*11*).

3.2.3 Computational studies on the inherent cyclization stereoselectivity

Computational simulations have been successfully applied previously to correlate inherent substrate reactivity with observed enzymatic selectivities (*19, 20*). To investigate the possible origins of the experimentally observed sequence-dependent stereochemical preferences, the cyclization reactions were studied computationally for the A rings of the $CylL_S$ and HalA2 core peptides (performed by Dr. Gonzalo Jiménez-Osés in Prof. K. N. Houk's laboratory, Department of Chemistry and Biochemistry, University of California, Los Angeles). The stereo-determining transition structures (TS) turned out to be highly diverse with respect to conformations presumably due to the high flexibility of the target oligopeptides. Therefore, a stepwise approach combining molecular dynamics (MD) and quantum mechanical (QM) calculations was employed to locate the optimal pathways for either *Re* or *Si* face additions for each model peptide.

The $CylL_S$ sequence leading to its A ring was found to prefer the LL stereochemistry by exhibiting a large difference in the activation barriers ($\Delta\Delta G^{\ddagger}$ = 3.2 kcal mol^{-1}) calculated for the Michael-type additions to the *Re* and *Si* faces of Dhb, corresponding to a stereoselectivity higher than 99% at room temperature (**Figure 3.12**). The A ring of HalA2 was also calculated to favor the *Re* face addition with activation energies that differed by 1.9 kcal mol^{-} (>96:4 *d.r.* at 25 ℃) (**Figure 3.13**), although the predicted secondary structures of the sequence is quite different from those calculated to be optimal for the A ring of $CylL_S$. The substitution of the 2^{nd} Dhb to Ala in the A ring of HalA2 resulted in minimal differences with respect to simulated structures comparing to the wild type A ring of HalA2. However, such a mutation had a strong impact on the calculated stereoselectivity, favoring the *Si* face approach by 2 kcal mol^{-1} (**Figure 3.14**). Collectively, computational studies of the intrinsic preferences for *Re* or *Si* face additions of three model peptide sequences corresponding to the A ring of $CylL_S$, and the wild type and mutated A rings of HalA2 are in accordance with the experimental observations, suggesting that such preferences for LL or DL stereochemistries are encoded in the substrate peptide sequences rather than dominated by the synthetases.

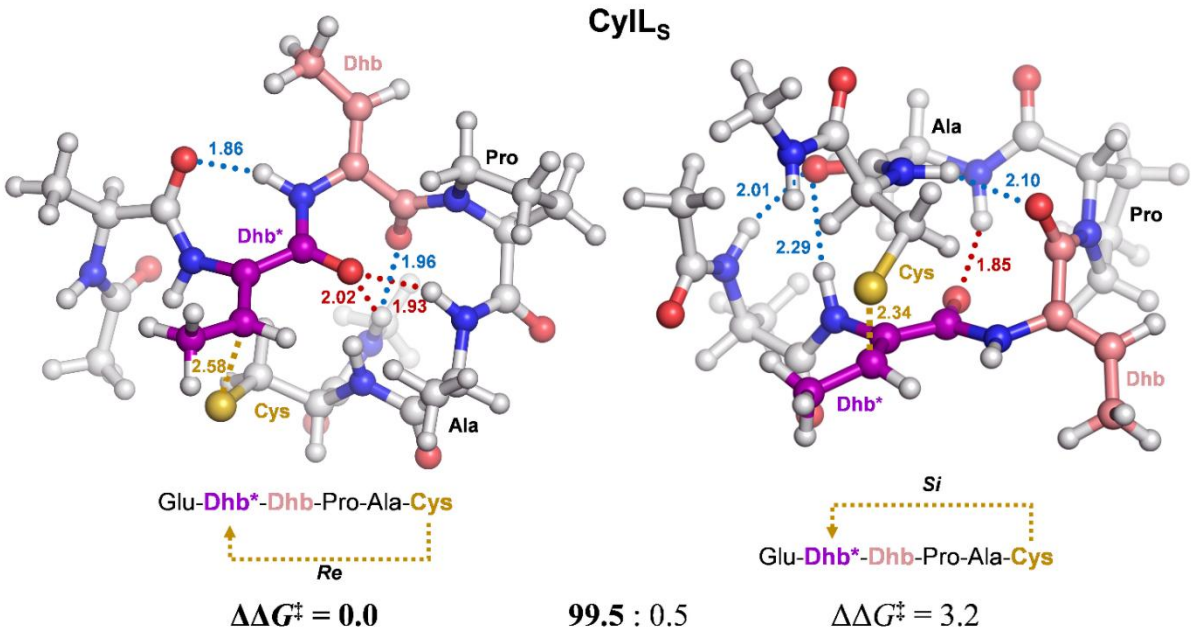

Figure 3.12 Lowest energy quantum mechanical transition structures for the *Re* and *Si* face Michael-type additions that generate the A ring in the CylL$_S$ core sequence. The depicted kinetic diastereomeric percentage was derived from these relative activation energies. Figure made by Dr. Gonzalo Jiménez-Osés.

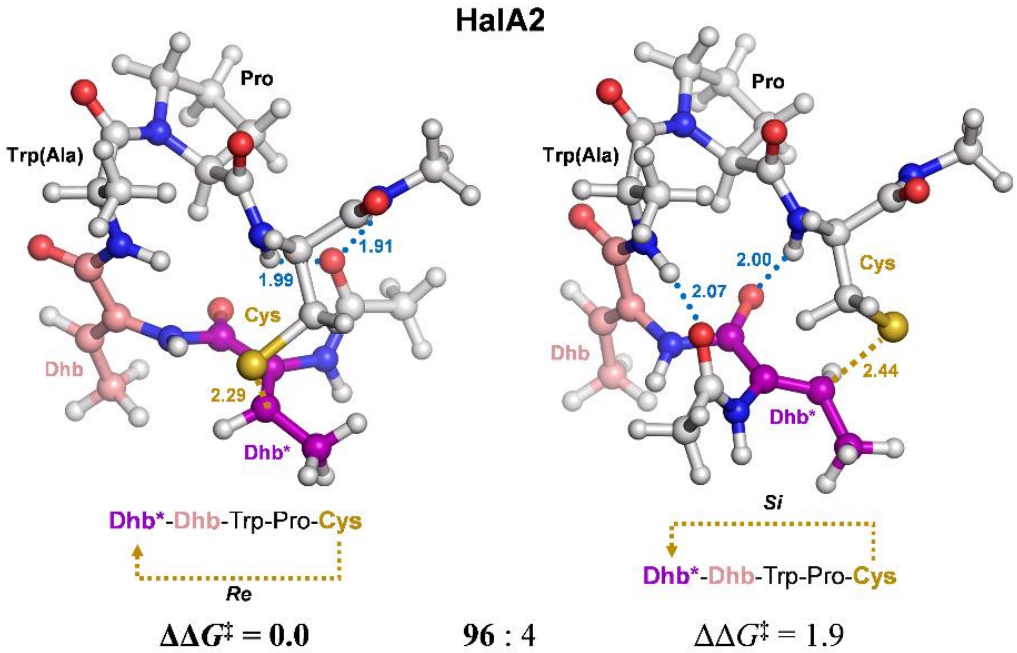

Figure 3.13 Lowest energy quantum mechanical transition structures for the *Re* and *Si* face Michael-type attack that generates the A ring in the HalA2 core sequence. The depicted kinetic diastereomeric percentages were derived from these relative activation energies. Figure made by Dr. Gonzalo Jiménez-Osés.

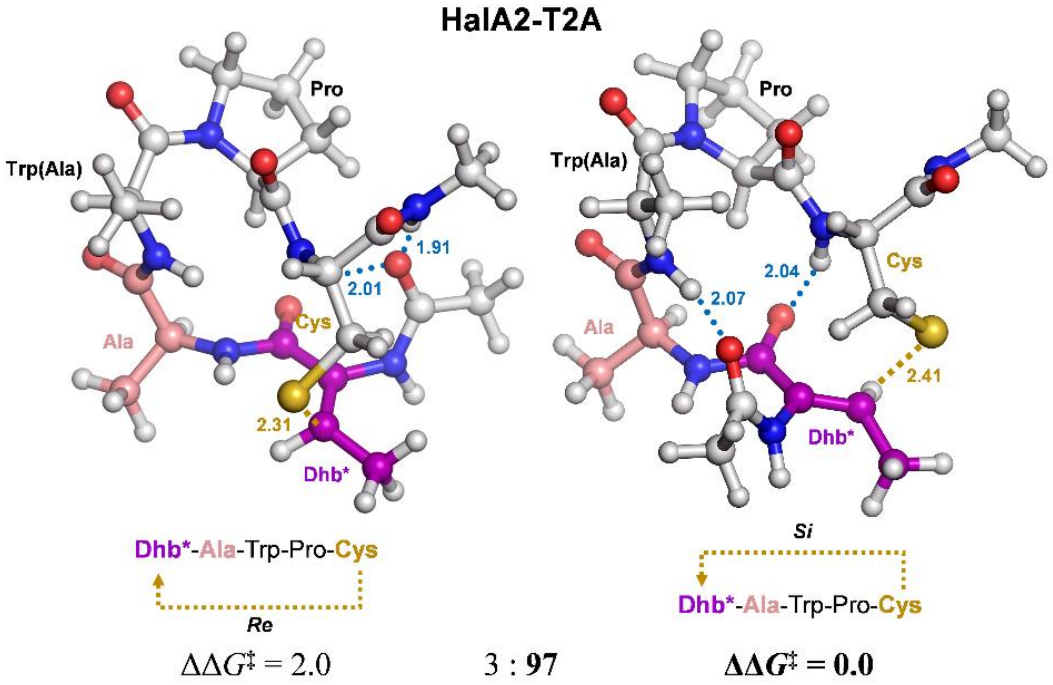

Figure 3.14 Lowest energy quantum mechanical transition structures for the *Re* and *Si* face Michael-type attack that generates the A ring of the HalA2-T2A core sequence. The depicted kinetic diastereomeric percentages were derived from these relative activation energies. Figure made by Dr. Gonzalo Jiménez-Osés.

Table 3.1 Stereochemical outcome of lanthipeptide precursor peptides modified by different lanthionine synthetases.

Peptide substrate	Lanthionine synthetase	Configuration of (Me)Lan residue		Quantum mechanical prediction for A ring
		A ring[a]	Other rings	
$CylL_S$	CylM[b]	LL	DL	LL
HalA2-$CylL_S$	HalM2	LL	DL	LL
LtnA2-$CylL_S$	LtnM2	LL	DL	LL
ProcA3.2-$CylL_S$	ProcM	LL	DL	LL
HalA2	HalM2[b]	LL	DL	LL
HalA2-T2A	HalM2	DL	DL	DL

a. The ring that contains the Dhb-Dhb-Xxx-Xxx-Cys motif in all substrates except HalA2-T2A which contain Dhb-Ala-Trp-Pro-Cys.

b. Reported previously (*3*).

3.3 Discussion

The discovery of LL-MeLan residues in the enterococcal cytolysin and haloduracin formed from a conserved Dhb-Dhb-Xxx-Xxx-Cys substrate motif revealed the first examples of enzymatic formation of non-canonical stereochemistry of MeLan residues (chapter 2). Very recently, LL stereochemistry was also reported for the lantibiotic carnolysin that also contains this motif (*21*). In this chapter, I show that the unusual stereochemistry is inherent to this sequence motif, because four different class II lanthionine synthetases with low sequence identities in their cyclase domains (<25% identity) all converted this motif to LL-MeLan products even though the enzymes install other rings in the same peptide with DL stereochemistry. Indeed, two of the tested lanthionine synthetases only give products with DL stereochemistry in their natural substrates, making it highly unlikely that they would have evolved features that favor LL-MeLan-containing products. Furthermore, the stereoselectivity was changed from LL to DL when the second Dhb in such a motif in HalA2 was mutated to Ala. Thus, these Michael-type additions are rare examples of natural enzymatic processes in which stereochemistry is substrate-controlled; other similar examples were not found by searching the literature, but a related observation has been made in ketoreduction by iterative polyketides synthases where substrate length controlled the face selectivity of hydride transfer (*22*). Of course, reversal of stereoselectivity of enzymatic reactions has been observed routinely with non-natural substrate analogs or mutant enzymes (*23*-*26*).

The computational studies also support a model in which the observed stereochemistry is determined to a large extent by the sequence of the substrate peptide, even though the transition state structures can differ substantially depending on the location of a Pro residue. Based on these observations, we suggest that for the formation of most (Me)Lan residues in lanthipeptides, the enzyme controls the face selectivity of attack and protonation, resulting in the predominantly observed DL stereochemistry. However, two consecutive Dhbs in the Dhb-Dhb-Xxx-Xxx-Cys motif may result in a conformational preference that is unfavorable for the canonical interaction with the enzyme active site. The enzyme is likely to still activate the Cys residue given the very slow rates of non-enzymatic (Me)Lan formation (*27*), but the face selectivity of attack is now governed by the substrate and not the enzyme. The role of the enzyme was investigated in more detail by decoupling the cyclization reaction from the dehydration reaction (see chapter 4). Structural information of substrate bound to the enzyme will be required to determine whether all calculated transition states can be attained in the context of the protein.

3.4 Methods

General methods and materials

Similar general methods and materials as what were described in chapter 2 were employed in this chapter unless specified otherwise. MALDI-TOF MS was carried out on a Bruker Daltonics UltrafleXtreme MALDI TOF/TOF instrument (Bruker). LC-ESI-Q/TOF MS and MS/MS analyses were conducted using a Synapt G1 instrument with an Acquity UPLC system (Waters), which was equipped with a Jupiter Proteo C12 column (5 μm; 90 Å; 100 x 1.0 mm) (Phenomenex). GC-MS analysis was performed in the Mass Spectrometry Laboratory (School of Chemical Sciences, UIUC) on an Agilent 7890 gas chromatograph (Agilent). Solid phase extraction was performed with Strata-X polymeric reverse phase columns (Phenomenex). Trypsin was purchased from Worthington Biochemical Corporation; Factor Xa was obtained from New England Biolabs and other endoproteinases were ordered from Roche Biosciences.

Construction of pRSFDuet-1 derivatives for co-expression of HalM2, LtnM2 and ProcM with HalA2-CylL$_S$, LtnA2-CylL$_S$ and ProcA3.2-CylL$_S$, respectively

Genes encoding the chimeric peptides were amplified by multi-step overlap extension PCR. First, the amplification of the 5' leader part was carried out by 30 cycles of denaturing (95 °C for 10 s), annealing (55 ℃ for 30 s), and extending (72 ℃ for 15 s) using appropriate primers (**Table 3.2**) to generate a forward megaprimer (FMP). In parallel, PCR reactions using appropriate forward primers (FP, **Table 3.2**) and the *cylL$_S$* reverse primer (RP) were also performed to produce 3' fragments (termed reverse megaprimer, RMP). The 5' FMP fragment and the 3' RMP fragment were purified by 2% agarose gel and combined in equal amounts (by mol, approximately 20 ng each for a 50 μL PCR reaction) and amplified using the same PCR conditions with *halA2*/*ltnA2*/*procA3.2* FP and *cylL$_S$* RP. The resulting PCR products were then cloned into the multiple cloning site 1 (MCS1) of pRSFDuet-1/HalM2-2(*28*), pRSFDuet-1/LtnM2-2 and pRSFDuet-1/ProcM-2 (*28*) vectors, respectively, to generate pRSFDuet-1/HalA2-CylL$_S$/HalM2-2, pRSFDuet-1/LtnA2-CylL$_S$/LtnM2-2, and pRSFDuet-1/ProcA3.2-CylL$_S$/ProcM-2. pRSFDuet-1/LtnM2-2 was constructed by amplifying the *ltnM2* gene out of the pBAC105 plasmid extracted from *Lactococcus lactis* IFPL105 (*29*) and cloning into the MCS 2 of a pRSFDuet-1 vector using *Nde*I and *Kpn*I restriction sites.

Construction of pRSFDuet-1 derivatives for co-expression of HalM2 and HalA2-T2A

The *halA2-T2A* gene was amplified from a pET15b/HalA2-T2A plasmid which was constructed by Dr. Leigh Anne Furgerson and cloned into MCS1 of the pRSFDuet-1/HalM2-2 vector using the *BamH*I and *Hind*III restriction sites to generate the pRSFDuet-1/HalA2-T2A/HalM2-2 plasmid. Primer sequences are listed in **Table 3.2**.

Table 3.2 Primer sequences for cloning of *halA2-cylL$_S$*, *ltnA2-cylL$_S$*, *procA3.2-cylL$_S$* and *halA2-T2A*.

Primer Name	Primer Sequence (5'-3')
HalA2_BamHI_FP	AAAAA GGATCC G ATGGTAAATT CAAAAGATTT
HalA2_HindIII_RP	AAAAA AAGCTT TTAGCACTGG CTTGTACACT
CylL$_S$_EcoRI_FP	AAAAA GAATTCG CTGAATA AAGAAAATCA G
CylL$_S$_NotI_RP	AAAAA GCGGCCGC TTAGCAA AATTTTGCGC T
CylL$_S$_HindIII_RP	AAAAA AAGCTT TTAGCAA AATTTTGCGC T
LtnM2_NdeI_FP	AAAAA CATATG TTGGACCCAT CAATAAAAAA A
LtnM2_KpnI_FP	AAAAA GGTACC TTAGGATTTA TGGCAACCAT C
LtnA2_BamHI_FP	AAAAA GGATCCG ATGAAAGAAA AAAATATGAA A
ProcA3.2_EcoRI_FP	AAAAA GAATTCG ATGTCAGAAGAACAACTCAAG
HalA2_Leader_RP	TTGTGCAT GCACATCTCC TGA
LtnA2-Leader_RP	TCCTCCATGA GACTCATCCC C
ProcA3.2_Leader_RP	TCC CCCAGCCACA CCTTCCAG
HalA2_CylL$_S$_FP	GATGTGCATGCACAAACCACACCGGCATGTTTTACC
LtnA2_CylL$_S$_FP	GAGTCTCATGGAGGAACCACACCGGCATGTTTTACC
ProcA3.2_CylL$_S$_FP	GGTGTGGCTGGGGGAACCACACCGGCATGTTTTACC

Expression and purification of modified His$_6$-HalA2-CylL$_S$, His$_6$-LtnA2-CylL$_S$, His$_6$-ProcA3.2-CylL$_S$ and His$_6$-HalA2-T2A peptides

E. coli BL21 (DE3) cells were transformed with the pRSFDuet-1/HalA2-CylL$_S$/HalM2-2, pRSFDuet-1/LtnA2-CylL$_S$/LtnM2-2, pRSFDuet-1/ProcA3.2-CylL$_S$/ProcM-2 or pRSFDuet-1/HalA2-T2A/HalM2-2 plamids and plated on an LB plate containing 50 mg/L kanamycin. A single colony was picked and grown in 10 mL of LB with 50 mg/L kanamycin aerobically at 37 ^oC for 10-16 h and the resulting culture was used to inoculate 1 L of LB. Cells were cultured at 37 ^oC until the OD at 600 nm reached 0.5, cooled and IPTG was added to a final concentration of 0.1 mM. The cells were cultured at 18 ^oC for another 18 h before harvesting. The cell pellet was resuspended at room temperature in LanA start buffer (20 mM NaH_2PO_4, pH 7.5 at 25 ^oC, 500 mM NaCl, 0.5 mM imidazole, 20% glycerol) and lysed by sonication. The sample was

centrifuged at 23,700 ×*g* for 30 min and the supernatant was kept. The pellet was then resuspended in LanA buffer 1 (6 M guanidine hydrochloride, 20 mM NaH_2PO_4, pH 7.5 at 25 ^oC, 500 mM NaCl, 0.5 mM imidazole) and sonicated again. The insoluble portion was removed by centrifugation at 23,700 ×*g* for 30 min and the soluble portion was kept. Both soluble portions were passed through 0.45-μm syringe filters and the His-tagged modified peptides were purified by immobilized metal affinity chromatography (IMAC) as previously described (*30*). The eluted fractions were desalted with Strata-X polymeric reverse phase SPE columns. The desalted peptides were lyophilized and stored at −20 ^oC for future use.

LC-ESI-Q/TOF MS and MS/MS analyses

A 5 μL volume of sample (approximately 100 μg/mL) was applied onto the column (Jupiter Proteo C12 column; 5 μm; 90 Å; 100 x 1.0 mm; Phenomenex) that was pre-equilibrated in aqueous solvent A. The solvents used for LC were: solvent A = 0.1% formic acid in water and solvent B = 0.1% formic acid in acetonitrile. A solvent gradient of 3%-97% B over 25 min was employed and the fractionated sample was directly subjected to ESI-Q/TOF MS analysis using a Waters Synapt G1 mass spectrometer. The mass spectrometer was calibrated before any sample was injected. Data were acquired in ESI positive mode with the capillary voltage set to 3.0-3.5 kV. The ionization source and desolvation gas were heated to 120 ℃ and 300 ℃, respectively. Cone gas was set to 150 L/h and desolvation gas was set to 600 L/h. The transfer collision energy was set to 4 V for both MS and MSMS analyses. The trap collision energy was set to 6 V for MS analysis. For MSMS analysis, a trap collision energy ramp ranging from 20-40 V was applied on multiply charged parent ions to achieve fragmentation. Suitable trap collision energy was determined by choosing the spectra where both fragment peaks and parent peak could be observed. [Glu^1]-Fibrinopeptide B (Sigma) was directly infused as lock mass with lock spray sampling. The acquired spectra were processed using MaxEnt3 software and analyzed by Protein/Peptide Editor in BioLynx 4.1 (Waters).

GC-MS analysis

The synthesis of Lan and MeLan standards and the preparation of samples for GC-MS analysis were carried out following a reported procedure published elsewhere (*10, 14*). The modified full length peptides with their leader peptides attached were hydrolyzed and the resulting

solutions were dried and directly used for derivatization. The derivatized samples were analyzed by GC-MS using an Agilent 7890 gas chromatograph equipped with a CP-Chirasil-L-Val fused silica column (25 m x 0.25 mm x 0.15 μm). Samples were dissolved in methanol and introduced to the instrument via a split (1:20) injection at a flow rate of 1.7 mL/min helium gas. The temperature method used was 160 ^{o}C for 2 min, 160 ^{o}C to 180 ^{o}C at 3 ^{o}C/min, and 180 ^{o}C for 2 min. The selected-ion monitoring (SIM) was achieved by extracting the total ion spectra for peaks containing the characteristic fragment masses of 365 Da for Lan and 379 Da for MeLan residues.

Protease cleavage of the leader peptides

Modified HalA2-CylL_S, ProcA3.2-CylL_S and HalA2-T2A peptides were dissolved in H_2O to a concentration of 3 mg/mL. To a total 200 μL reaction with a final peptide concentration of 200 μM, 20 μL of 500 mM HEPES buffer (pH 7.5) was added, followed by 10 μL of 0.02 mg/mL AspN protease (for HalA2-CylL_S), 1 mg/mL GluC protease (for ProcA3.2-CylL_S), and 1 mg/mL Factor Xa or 5 mg/mL trypsin (for HalA2-T2A). The protease cleavage reactions were incubated at 25 ^{o}C for 6 to 16 h.

Computational simulations

All MD simulations and DFT calculations were performed by Dr. Gonzalo Jiménez-Osés and Dr. K. N. Houk. For detailed procedures, please refer to Tang, W., Jimenez-Oses, G., Houk, K. N., and van der Donk, W. A. (2015) Substrate control in stereoselective lanthionine biosynthesis, *Nat. Chem. 7*, 57-64.

3.5 References

1. Masamune, S., Choy, W., Petersen, J. S., and Sita, L. R. (1985) Double asymmetric synthesis and a new strategy for stereochemical control in organic synthesis, *Angew. Chem. Int. Ed. Engl. 24*, 1-30.
2. Chatterjee, C., Paul, M., Xie, L., and van der Donk, W. A. (2005) Biosynthesis and mode of action of lantibiotics, *Chem. Rev. 105*, 633-684.
3. Tang, W., and van der Donk, W. A. (2013) The sequence of the enterococcal cytolysin imparts unusual lanthionine stereochemistry, *Nat. Chem. Biol. 9*, 157-159.
4. Cox, C. R., Coburn, P. S., and Gilmore, M. S. (2005) Enterococcal cytolysin: a novel two component peptide system that serves as a bacterial defense against eukaryotic and prokaryotic cells, *Curr. Prot. Pept. Sci. 6*, 77-84.
5. Huycke, M. M., Spiegel, C. A., and Gilmore, M. S. (1991) Bacteremia caused by hemolytic, high-level gentamicin-resistant *Enterococcus faecalis*, *Antimicrob. Agents Chemother. 35*, 1626-1634.
6. Libertin, C. R., Dumitru, R., and Stein, D. S. (1992) The hemolysin/bacteriocin produced by enterococci is a marker of pathogenicity, *Diagn. Microbiol. Infect. Dis. 15*, 115-120.
7. Chow, J. W., Thal, L. A., Perri, M. B., Vazquez, J. A., Donabedian, S. M., Clewell, D. B., and Zervos, M. J. (1993) Plasmid-associated hemolysin and aggregation substance production contribute to virulence in experimental enterococcal endocarditis, *Antimicrob. Agents Chemother. 37*, 2474-2477.
8. McClerren, A. L., Cooper, L. E., Quan, C., Thomas, P. M., Kelleher, N. L., and van der Donk, W. A. (2006) Discovery and in vitro biosynthesis of haloduracin, a two-component lantibiotic, *Proc. Natl. Acad. Sci. U.S.A. 103*, 17243-17248.
9. Shi, Y., Yang, X., Garg, N., and van der Donk, W. A. (2011) Production of lantipeptides in *Escherichia coli*, *J. Am. Chem. Soc. 133*, 2338-2341.
10. Liu, W., Chan, A. S., Liu, H., Cochrane, S. A., and Vederas, J. C. (2011) Solid supported chemical syntheses of both components of the lantibiotic lacticin 3147, *J. Am. Chem. Soc. 133*, 14216-14219.
11. Kido, Y., Hamakado, T., Yoshida, T., Anno, M., Motoki, Y., Wakamiya, T., and Shiba, T. (1983) Isolation and characterization of ancovenin, a new inhibitor of angiotensin I converting enzyme, produced by actinomycetes, *J. Antibiot. 36*, 1295-1299.
12. Suda, S., Cotter, P. D., Hill, C., and Ross, R. P. (2012) Lacticin 3147- biosynthesis, molecular analysis, immunity, bioengineering and applications, *Curr. Protein Pept. Sci. 13*, 193-204.
13. Li, B., Sher, D., Kelly, L., Shi, Y. X., Huang, K., Knerr, P. J., Joewono, I., Rusch, D., Chisholm, S. W., and van der Donk, W. A. (2010) Catalytic promiscuity in the biosynthesis of cyclic peptide secondary metabolites in planktonic marine cyanobacteria, *Proc. Natl. Acad. Sci. U.S.A. 107*, 10430-10435.

14. Tang, W., and van der Donk, W. A. (2012) Structural characterization of four prochlorosins: a novel class of lantipeptides produced by planktonic marine cyanobacteria, *Biochemistry 51*, 4271-4279.
15. Zhang, Q., Yu, Y., Velásquez, J. E., and van der Donk, W. A. (2012) Evolution of lanthipeptide synthetases, *Proc. Natl. Acad. Sci. U. S. A. 109*, 18361-18366.
16. Zhang, Q., Yang, X., Wang, H., and van der Donk, W. A. (2014) High divergence of the precursor peptides in combinatorial lanthipeptide biosynthesis, *ACS Chem. Biol. 9*, 2686-2694.
17. Bindman, N. A., and van der Donk, W. A. (2013) A general method for fluorescent labeling of the N-termini of lanthipeptides and its application to visualize their cellular localization, *J. Am. Chem. Soc. 135*, 10362-10371.
18. Garg, N., Tang, W., Goto, Y., Nair, S. K., and van der Donk, W. A. (2012) Lantibiotics from *Geobacillus thermodenitrificans*, *Proc. Natl. Acad. Sci. U.S.A. 109*, 5241-5246.
19. Michelson, A. Z., Rozenberg, A., Tian, Y., Sun, X., Davis, J., Francis, A. W., O'Shea, V. L., Halasyam, M., Manlove, A. H., David, S. S., and Lee, J. K. (2012) Gas-phase studies of substrates for the DNA mismatch repair enzyme MutY, *J. Am. Chem. Soc. 134*, 19839-19850.
20. Zu, L., Xu, M., Lodewyk, M. W., Cane, D. E., Peters, R. J., and Tantillo, D. J. (2012) Effect of isotopically sensitive branching on product distribution for pentalenene synthase: support for a mechanism predicted by quantum chemistry, *J. Am. Chem. Soc. 134*, 11369-11371.
21. Lohans, C. T., Li, J. L., and Vederas, J. C. (2014) Structure and biosynthesis of carnolysin, a homologue of enterococcal cytolysin with D-amino acids, *J. Am. Chem. Soc. 136*, 13150-13153.
22. Zhou, H., Gao, Z., Qiao, K., Wang, J., Vederas, J. C., and Tang, Y. (2012) A fungal ketoreductase domain that displays substrate-dependent stereospecificity, *Nat. Chem. Biol. 8*, 331-333.
23. May, O., Nguyen, P. T., and Arnold, F. H. (2000) Inverting enantioselectivity by directed evolution of hydantoinase for improved production of L-methionine, *Nat. Biotechnol. 18*, 317-320.
24. Turner, N. J. (2003) Controlling chirality, *Curr. Opin. Biotechnol. 14*, 401-406.
25. Mugford, P. F., Wagner, U. G., Jiang, Y., Faber, K., and Kazlauskas, R. J. (2008) Enantiocomplementary enzymes: classification, molecular basis for their enantiopreference, and prospects for mirror-image biotransformations, *Angew. Chem. Int. Ed. 47*, 8782-8793.
26. Reetz, M. T. (2011) Laboratory evolution of stereoselective enzymes: a prolific source of catalysts for asymmetric reactions, *Angew. Chem. Int. Ed. 50*, 138-174.
27. Mukherjee, S., and van der Donk, W. A. (2014) Mechanistic studies on the substrate-tolerant lanthipeptide synthetase ProcM, *J. Am. Chem. Soc. 136*, 10450-10459.

28. Shi, Y., Yang, X., Garg, N., and van der Donk, W. A. (2011) Production of lantipeptides in *Escherichia coli*, *J. Am. Chem. Soc. 133*, 2338-2341.
29. Garcia-Ruiz, A., Requena, T., Pelaez, C., Bartolome, B., Moreno-Arribas, M. V., and Martinez-Cuesta, M. C. (2013) Antimicrobial activity of lacticin 3147 against oenological lactic acid bacteria. Combined effect with other antimicrobial agents, *Food Control 32*, 477-483.
30. Li, B., Cooper, L. E., and van der Donk, W. A. (2009) *In vitro* studies of lantibiotic biosynthesis *Methods Enzymol. 458*, 533-558.

Chapter 4. The Enterococcal Cytolysin Synthetase Coevolves with Substrate for Stereoselective Lanthionine Synthesis

4.1 Introduction

As a major family of natural products, RiPPs (chapter 1) are structurally diverse as a consequence of a multitude of different post-translational modifications, including disulfide formation, methylation, hydroxylation, decarboxylation, cyclodehydration, macrocyclization (*1, 2*). Of all the structural elements, one particular feature that accounts significantly for diversity and complexity in other natural products, the stereochemistry, has not received much attention in research related to RiPPs. The de-emphasis of stereochemistry is in part because RiPPs are initially assembled by the ribosome with amino acid building blocks that always adopt the L configuration. However, new stereocenters may be introduced and existing stereocenters can be modified during the post-translational modification process (*3-5*). For instance, the presence of 18 epimerized amino acids that adopt the non-canonical D configuration in polytheonamide, a RiPP of 48 amino acid residues, reinforces the idea that stereocenters inherited from the amino acid building blocks of the precursor peptide can be altered in the final natural product (*3*). Introduction of new stereocenters on side chains of selected amino acids as post-translational modifications can further increase the structural diversity of RiPPs. Thus far, little is known how these stereoselective modifications are achieved during RiPP synthesis.

Lanthipeptides are a family of RiPPs that have been intensively studied over the past 40 years (**Figure 1.2**) (*6, 7*). For an introduction of lanthipeptides and class II lanthionine synthetases, please chapters 1 and 3. The formation of dehydroamino acids as intermediates during lanthipeptide biosynthesis allows the possible generation of non-canonical one (Ser) or two (Thr) stereocenters from the amino acid building blocks in the final Lan- or MeLan-containing products. Based on the characterization of a handful of members in the lanthipeptide family, it was generally assumed that the original L stereocenter of these residues is inverted to D during the Michael-type addition to form a DL stereochemistry in the resulting thioether crosslinked moiety (i.e. D configuration for the α carbon of the former Ser/Thr residue and L configuration for the α carbon of the former Cys residue; (2*S*, 6*R*) for Lan and (2*S*, 3*S*, 6*R*) for MeLan, **Figure 2.6a**, left panel) (*8*). As described in chapter 2, this assumption was only recently

challenged by the discovery of an unusual LL stereochemistry in the enterococcal cytolysin (i.e. L configuration for both α carbons; (2*R*, 6*R*) for Lan and (2*R*, 3*R*, 6*R*) for MeLan, **Figure 2.6a**, right panel) (*9*). The enterococcal cytolysin is produced by clinical isolates of *Enterococcus faecalis* and is made up of two post-translationally modified peptides named cytolysin L and S (chapter 2) (*10*). Due to its tight linkage with the virulence of *E. faecalis*, cytolysin has been intensively studied since the 1930s with respect to its biosynthesis and biological activities (*11-15*). However, structural characterization of this compound revealed the existence of an unexpected stereochemistry of (Me)Lan residues that differs from all other lanthipeptides with documented structural information. Further investigation showed that the stereoselectivity is encoded in the substrate sequence rather than being governed by the enzyme, which is rare for naturally occurring enzymatic reactions (*16*). These discoveries bring to light the complexity of stereochemical control for RiPP biosynthesis and emphasize the importance of stereochemical characterization for unknown or even known RiPPs.

Although it has been confirmed that the unusual LL stereochemistry in cytolysin is induced by the substrate sequence, the role that the cytolysin synthetase CylM plays during this process has thus far not been investigated. In this chapter, I provide evidence that CylM has coevolved with its substrate to enforce the desired stereoselectivity for lanthionine synthesis in cytolysin. Moreover, the dehydrated cytolysin peptide seems to adopt a pre-organized conformation that facilitates the formation of the desired cyclized product even without enzymatic catalysis, reinforcing the idea of substrate-controlled reactivity and selectivity in RiPP biosynthesis.

4.2 Results

Chapter 3 described that three different lanthionine synthetases sharing limited sequence homology with CylM in their cyclase domain (<25% identity) all formed the A ring in $CylL_S$ with the LL stereochemistry that is found in the natural product (*16*). In that study, the dehydration of the $CylL_S$ peptide was executed in tandem with the cyclization reaction in *E. coli*, which prevents the direct investigation of the Michael-type addition process. In order to study the cyclization reaction in isolation, I first targeted obtaining dehydrated $CylL_S$ peptide without the formation of thioether linkages. LanM proteins are bifunctional enzymes that catalyze both the dehydration and cyclization reactions in the substrate peptides and these two functions are executed by two distinct domains- an N-terminal dehydratase domain and a C-terminal cyclase

domain (*17*, *18*). The crystal structure of CylM reveals that each domain has its own independent active site (chapter 5) (*19*). In addition, the activity of the dehydratase domain of one LanM protein, the lacticin 481 synthetase LctM, has recently been reconstituted *in vitro* (*20*), indicating that the dehydratase domain of LanM proteins can be catalytically competent without the partner cyclase domain. This observation provided a possible route to obtain dehydrated $CylL_S$ using truncated CylM. Therefore, a truncated CylM consisting of residues 1 to 625 that encompass the dehydratase domain was overexpressed and used to modify the substrate $CylL_S$ in *E. coli*. The peptide was then purified using immobilized metal affinity chromatography. Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) analysis illustrated a mass shift corresponding to loss of four water molecules, consistent with what was observed for fully dehydrated $CylL_S$ (**Figure 4.1**). The desired mass loss indicates that the CylM dehydratase domain is functional in the absence of the cyclase domain. To test whether the two Cys residues in $CylL_S$ obtained using this strategy were cyclized, N-ethylmaleimide (NEM) was employed as an alkylation regent to detect free thiols in the peptide (*21*). Unexpectedly, dehydrated $CylL_S$ did not react with NEM under reducing conditions (**Figure 4.2**), suggesting the thiols of both Cys5 and Cys21 were involved in covalent linkages other than disulfide bonds. Removal of the leader peptide by CylA, the dedicated serine protease encoded in the cytolysin biosynthetic pathway (chapter 7), allowed liquid chromatography-MS (LC-MS) analysis of the modified core peptide. One major peak with a mass corresponding to fully dehydrated $CylL_S$ was observed on the LC chromatogram and the retention time matched that of the cytolysin S peptide modified by full length CylM (**Figures 4.3a** and **4.4a**). Notably, a minor peak with the same mass but a different retention time was also detected on the LC chromatogram (**Figure 4.3a**). Further MS/MS analyses of both peaks revealed a similar fragmentation pattern as wild type cytolysin S (**Figures 4.3b** and **4.4b**), indicating that $CylL_S$ peptide obtained in this way was already in its cyclized form and that the regioselectivity of cyclization was the same as that in the CylM-catalyzed reaction. The minor product could be a diastereomer of wild type cytolysin S. These observations were quite surprising as they indicate that regioselective cyclization of $CylL_S$ could be achieved without the cyclase domain, and may suggest that dehydrated $CylL_S$ attains an organized conformation that auto-cyclizes even without the presence of a catalyst.

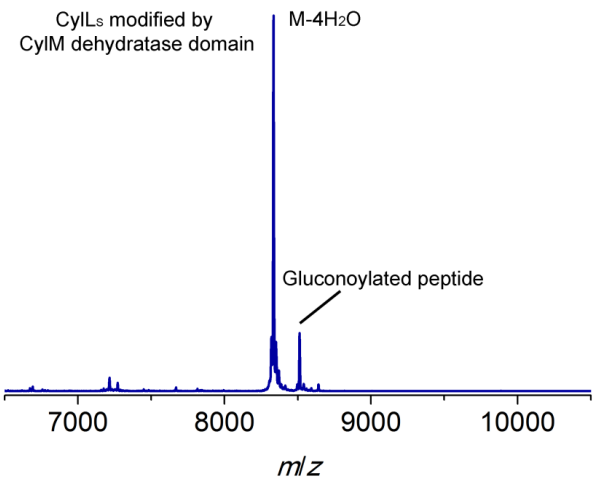

Figure 4.1 MALDI-TOF mass spectrum for $CylL_S$ modified by the CylM dehydratase domain in *E. coli*. Calculated $M–4H_2O$: 8,330, average mass; observed $M–4H_2O+H^+$: 8,334, average mass. Partial gluconoylation at the N terminus of dehydrated $CylL_S$ occurred when expressing the peptide in *E. coli* BL21(DE3), resulting in a +178 Da peak in addition to the desired peptide mass (*22*).

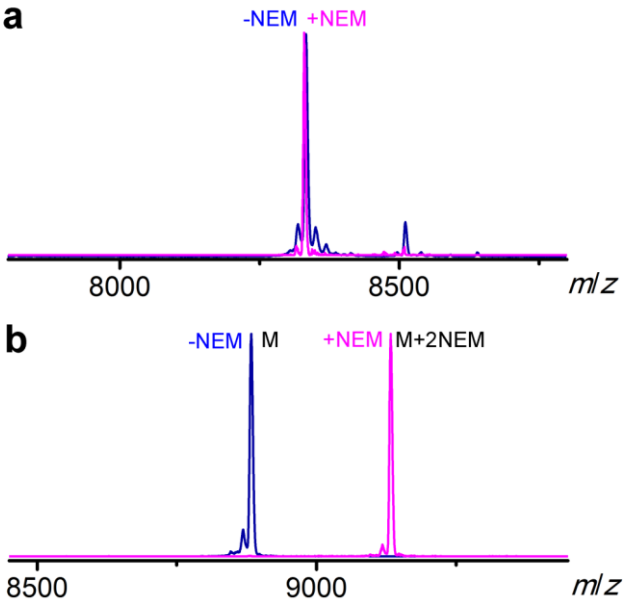

Figure 4.2 NEM analysis of $CylL_S$ modified by the CylM dehydratase domain in *E. coli*. Overlaid mass spectra of dehydrated $CylL_S$ (**a**) (blue trace) and the peptide treated with NEM (magenta trace). Unmodified $CylL_S$ (**b**) was employed as a control. Two NEM adducts were observed for unmodified $CylL_S$ that contains two free thiols, whereas no NEM adducts were detected for dehydrated $CylL_S$, suggesting that both thiols in this peptide were involved in thioether crosslinks. For dehydrated $CylL_S$, calculated $M–4H_2O$: 8,330, average mass; observed $M–4H_2O+H^+$: 8,334, average mass; for dehydrated $CylL_S$ treated by NEM, calculated $M–4H_2O+0NEM$: 8,330, average mass; observed $M–4H_2O+0NEM+H^+$: 8,330, average mass. For unmodifed $CylL_S$, calculated M: 8,883, average mass; observed $M+H^+$: 8,883, average mass; for unmodifed $CylL_S$ treated by NEM, calculated M+2NEM: 9,133, average mass; observed $M+2NEM+H^+$: 9,133, average mass.

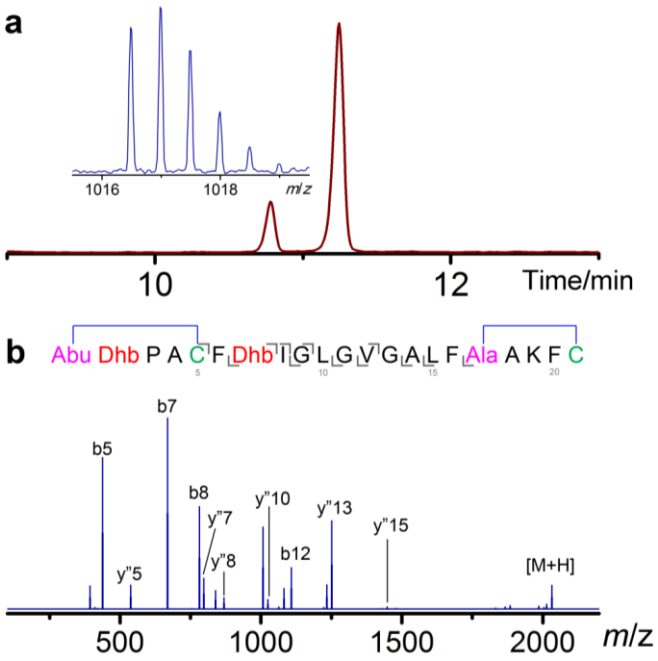

Figure 4.3 MS analyses of cytolysin S peptide modified by the CylM dehydratase domain in *E. coli*. (**a**) Extracted ion chromatogram (*m*/*z* = 1017) of $CylL_S$ core peptide modified by the CylM dehydratase domain. The ESI mass spectrum corresponding to the major peak is shown in the insert. For fully modified $CylL_S$ core peptide, calculated $[M–4H_2O+2H]^{2+}$: 1,016.52, monoisotopic mass; observed $[M–4H_2O+2H]^{2+}$: 1,016.51, monoisotopic mass. (**b**) ESI MS/MS analysis of the major peak on the LC chromatogram and a proposed structure deduced from the fragmentation pattern. Both MS and MS/MS results suggest that the dehydrated $CylL_S$ purified from *E. coli* is in its cyclized form.

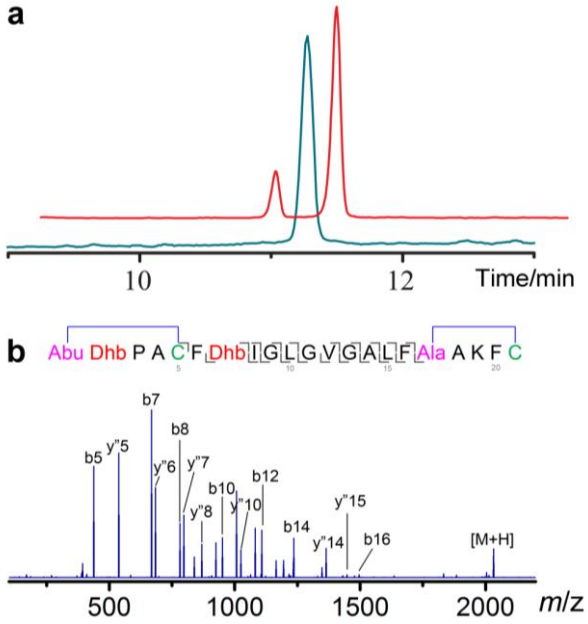

Figure 4.4 MS analyses of *in vivo* non-enzymatically cyclized cytolysin S. (**a**) Overlaid extracted ion chromatograms (*m*/*z* = 1017) of the $CylL_S$ core peptides modified by CylM (cyan trace) or the CylM dehydratase domain (red trace) in *E. coli*. (**b**) ESI MS/MS analysis of the minor product of *in vivo* cyclized cytolysin S and a proposed structure deduced from the fragmentation pattern.

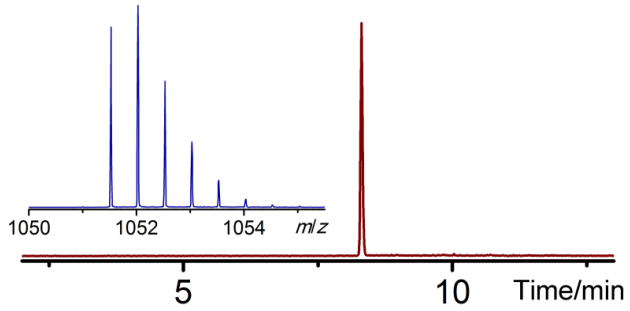

Figure 4.5 ESI-LC-MS analysis of linear CylL$_S$ peptide in its oxidized form. Extracted ion chromatogram (*m*/*z* = 1052) of the LC trace for oxidized CylL$_S$ core peptide is shown in dark red and the corresponding mass spectrum is shown in the insert. Linear CylL$_S$ peptide was expressed from *E. coli* and purified in its oxidized form. No peak corresponding to reduced CylL$_S$ core peptide was detected in the LC trace. Oxidized CylL$_S$ core peptide with a disulfide linkage, calculated $[M+2H]^{2+}$: 1,051.53, monoisotopic mass; observed $[M+2H]^{2+}$: 1,051.53, monoisotopic mass.

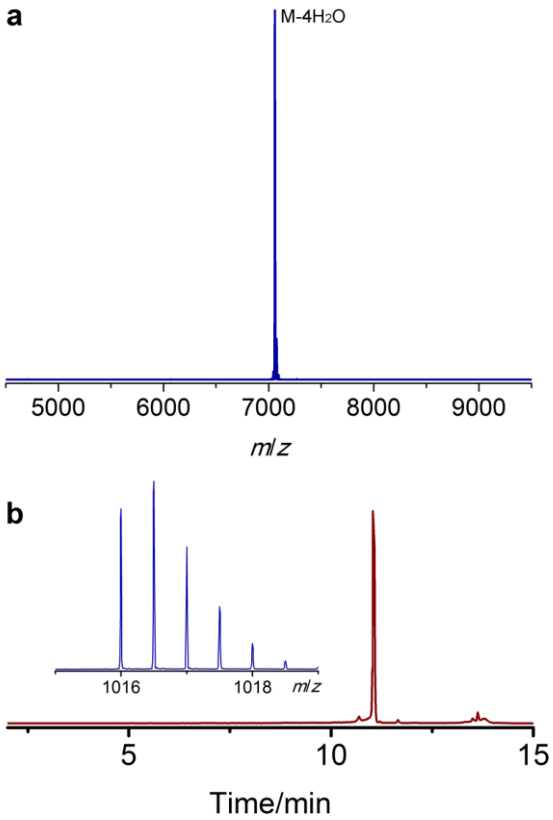

Figure 4.6 MS analyses of oxidized and dehydrated CylL$_S$ peptide. (**a**) MALDI-TOF mass spectrum of oxidized CylL$_S$ modified by CylM. For oxidized and dehydrated CylL$_S$ peptide, calculated M–4H$_2$O: 7,058, average mass; observed M–4H$_2$O+H$^+$: 7,059, average mass. (**b**) Extracted ion chromatogram (ESI, *m*/*z* = 1016.5) of the LC trace for oxidized and dehydrated CylL$_S$ core peptide with the corresponding mass spectrum inserted. Dhb1 was hydrolyzed to a ketone after leader peptide removal, resulting in a mass increase of 1 Da. Oxidized and dehydrated CylL$_S$ core peptide, calculated $[M–4H_2O–NH+O+2H]^{2+}$: 1,016.00, monoisotopic mass; observed $[M–4H_2O–NH+O+2H]^{2+}$: 1,016.00, monoisotopic mass.

To rule out the possibility that some *E. coli* proteins facilitated the Michael-type addition, the cyclization reaction of dehydrated $CylL_S$ was conducted *in vitro*. The two cysteines in $CylL_S$ were first protected as an intramolecular disulfide, and the oxidized $CylL_S$ was supplied to CylM (**Figure 4.5**). As no free thiols were available, cyclization could not take place during the process. CylM accepted oxidized $CylL_S$ as substrate and eliminated four water molecules (**Figure 4.6**), albeit with a lower efficiency compared to linear $CylL_S$. The 4-fold dehydrated $CylL_S$ peptide was then purified by reversed phase high performance liquid chromatography (RP-HPLC) (**Figure 4.7**). Subsequent incubation with dithiothreitol at pH 7.5 released the thiols and allowed the non-enzymatic cyclization to proceed. Free thiols could no longer be detected after 12 hours of incubation at room temperature (**Figure 4.8**). Further treating the peptide with the peptidase CylA and analysis of the core peptide using LC-MS again revealed a major peak on the LC chromatogram corresponding to fully cyclized $CylL_S$ as well as a minor peak with the same mass (**Figure 4.9a**). ESI MS/MS analyses of both products demonstrated the formation of the correct ring topology (**Figures 4.9b** and **4.9c**). As a result, the *in vitro* results are in good agreement with the *in vivo* observations, suggesting that the dehydrated $CylL_S$ cyclizes non-enzymatically into the correct ring topology.

I next examined the stereochemistry of (Me)Lan residues in non-enzymatically cyclized $CylL_S$ that was obtained from *E. coli*. A similar GC-based method was employed as described in chapter 2. Signals corresponding to LL-MeLan and DL-Lan were predominantly observed (**Figure 4.10**), suggesting that stereoselective cyclization of dehydrated $CylL_S$ could be achieved non-enzymatically. A notable LL-Lan peak was also detected in the GC trace, which is not observed for CylM-cyclized $CylL_S$ (*9*). This particular LL diastereomer of Lan was attributed to a diastereomer of cytolysin S containing an LL-MeLan in the A ring and an LL-Lan in the B ring, which corresponds to the minor peak on the LC chromatogram of non-enzymatically cyclized $CylL_S$ core peptide (**Figures 4.3a** and **4.9a**).

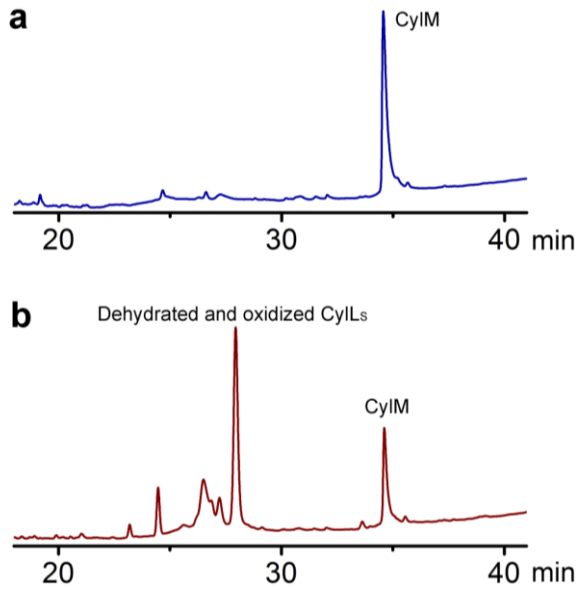

Figure 4.7 HPLC traces of CylM (a) and oxidized $CylL_S$ modified by CylM (b).

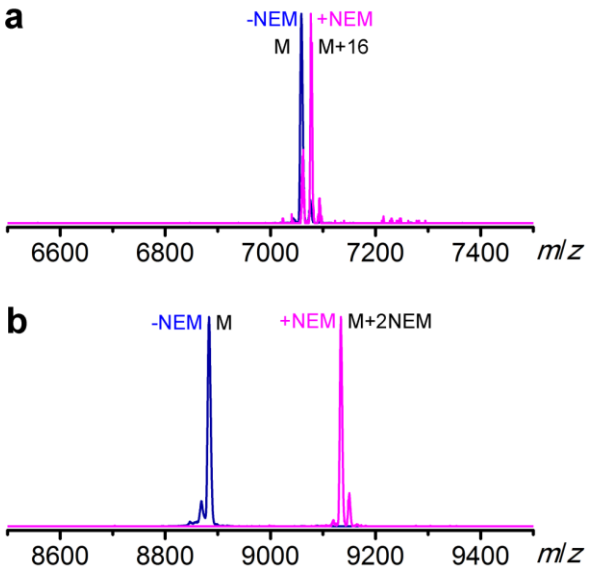

Figure 4.8 NEM analysis of *in vitro* non-enzymatically cyclized $CylL_S$. Overlaid MALDI-TOF mass spectra of dehydrated $CylL_S$ (**a**) (blue trace) and non-enzymatically cyclized $CylL_S$ peptide treated with NEM (magenta trace). Unmodified $CylL_S$ (**b**) was employed as a control. Two NEM adducts were observed for unmodified $CylL_S$ that contains two free thiols, whereas no NEM adducts were detected for non-enzymatically cyclized $CylL_S$. A large +16 peak was observed after 12 hours of incubation at room temperature for non-enzymatically cyclized $CylL_S$, which was attributed to oxidation mainly in the leader peptide region, as little such adduct was detected when analyzing the core peptide (see **Figure 4.9**). For oxidized and dehydrated $CylL_S$, calculated M–4H_2O: 7,058, average mass; observed M–4H_2O+H^+: 7,059, average mass; for non-enzymatically cyclized $CylL_S$ treated by NEM, calculated M–4H_2O+0NEM: 7,060, average mass; observed M–4H_2O+0NEM+H^+: 7,062 and 7,078 (possibly M–4H_2O+0NEM+O+H^+), average mass. For unmodifed $CylL_S$, calculated M: 8,883, average mass; observed M+H^+: 8,883, average mass; for unmodifed $CylL_S$ treated by NEM, calculated M+2NEM: 9,133, average mass; observed M+2NEM+H^+: 9,134, average mass.

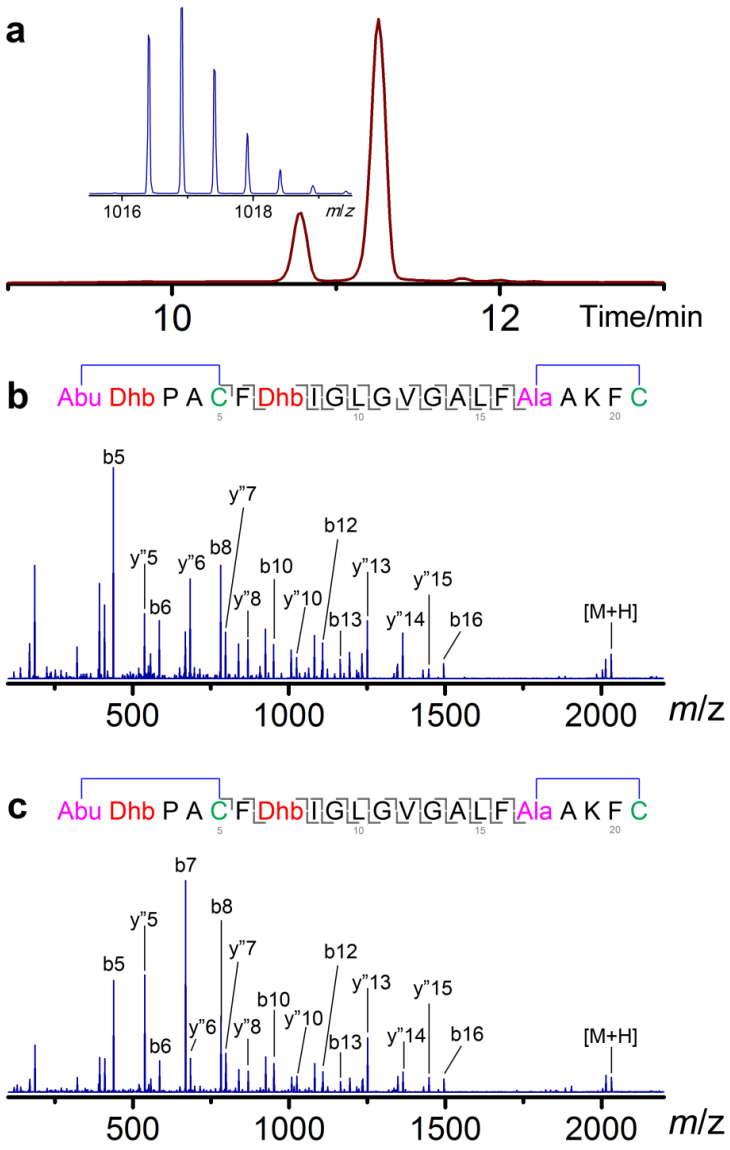

Figure 4.9 MS analyses of *in vitro* non-enzymatically cyclized cytolysin S peptide. (**a**) Extracted ion chromatogram (*m*/*z* = 1017) of the LC trace for *in vitro* non-enzymatically cyclized $CylL_S$ core peptide. The mass spectrum corresponding to the major peak is shown in the insert. For modified cytolysin S peptide, calculated $[M–4H_2O+2H]^{2+}$: 1,016.52, monoisotopic mass; observed $[M–4H_2O+2H]^{2+}$: 1,016.40, monoisotopic mass. (**b**) ESI MS/MS analysis of the major product of non-enzymatically cyclized cytolysin S and a proposed structure deduced from the fragmentation pattern. (**c**) ESI MS/MS analysis of the minor product of non-enzymatically cyclized cytolysin S and a proposed structure deduced from the fragmentation pattern.

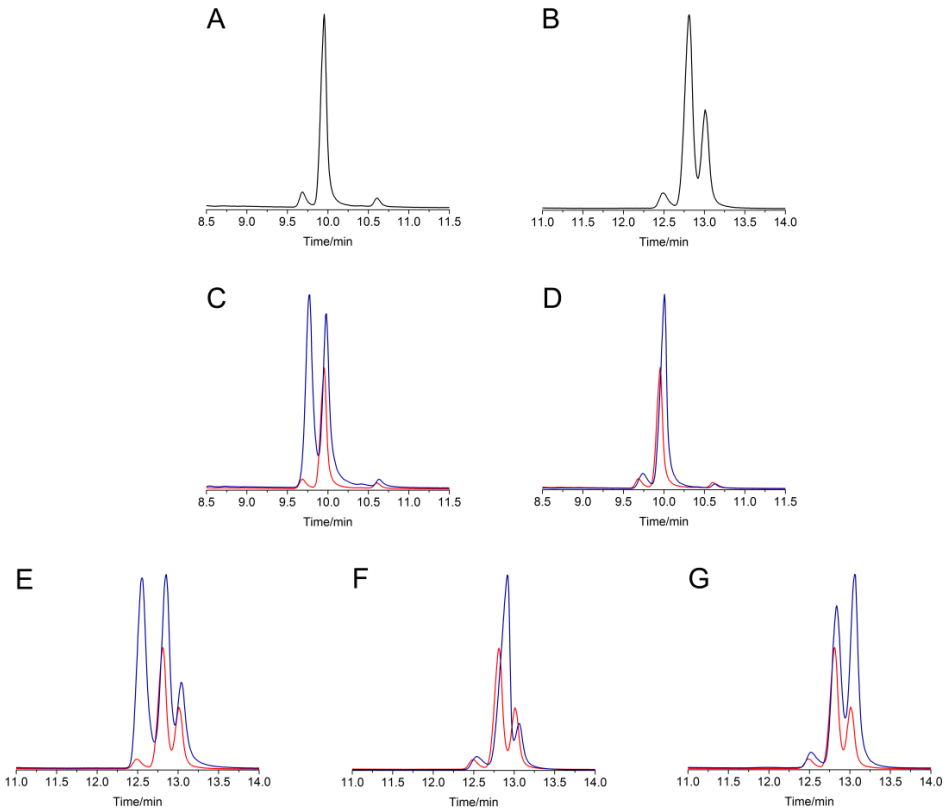

Figure 4.10 GC-MS traces for hydrolyzed/derivatized Lan/MeLan residues obtained from non-enzymatically cyclized CylL$_S$ in *E. coli* and co-injections with derivatized Lan/MeLan standards (selected ion monitoring, SIM, at 365 Da for Lan and 379 Da for MeLan). **(a)** Hydrolyzed and derivatized MeLan residues from non-enzymatically cyclized CylL$_S$. **(b)** Hydrolyzed and derivatized Lan residues from non-enzymatically cyclized CylL$_S$. **(c)** Hydrolyzed and derivatized MeLan residues from non-enzymatically cyclized CylL$_S$ (red line) and co-injected with derivatized DL-MeLan standard (blue line). **(d)** Hydrolyzed and derivatized MeLan residues from non-enzymatically cyclized CylL$_S$ (red line) and co-injected with derivatized LL-MeLan standard (blue line). **(e)** Hydrolyzed and derivatized Lan residues from non-enzymatically cyclized CylL$_S$ (red line) and co-injected with derivatized DD-Lan standard (blue line). **(f)** Hydrolyzed and derivatized Lan residues from non-enzymatically cyclized CylL$_S$ (red line) and co-injected with derivatized DL-Lan standard (blue line). **(g)** Hydrolyzed and derivatized Lan residues from non-enzymatically cyclized CylL$_S$ (red line) and co-injected with derivatized LL-Lan standard (blue line). Traces of derivatized Lan and MeLan residues from non-enzymatically cyclized CylL$_S$ (red lines) used for overlay were adjusted to 70% intensity for clarity.

The observations that dehydrated $CylL_S$ cyclizes mainly into the correct product without any cyclase raised the question whether the CylM cyclase domain is required for cytolysin S synthesis. However, both *in vitro* and *in vivo* non-enzymatic cyclization resulted in a minor product corresponding to a diastereomer of cytolysin S, where the configuration of the Lan residue in the B ring was inverted from DL to LL. Therefore, I wondered whether the CylM cyclase domain could improve the fidelity for the cyclization of the B ring in $CylL_S$. To test this hypothesis, the CylM cyclase domain consisting of residues 626 to 993 and the CylM dehydratase domain were expressed together with $CylL_S$ and allowed to modify the peptide *in trans* in *E. coli*. Fully dehydrated and cyclized product was obtained (**Figures 4.11** and **4.12**), and indeed, the minor peak corresponding to the diastereomer of cytolysin S was greatly diminished in intensity when analyzing the core peptide by LC-MS (**Figure 4.13**). The LL-Lan signal on the GC chromatogram also returned to a basal level (**Figure 4.14**), resulting in a similar GC trace as what was observed for CylM-modified $CylL_S$. Collectively, these results suggest a role of the CylM cyclase domain in CylM-catalyzed modifications with respect to the stereochemical fidelity of Lan formation.

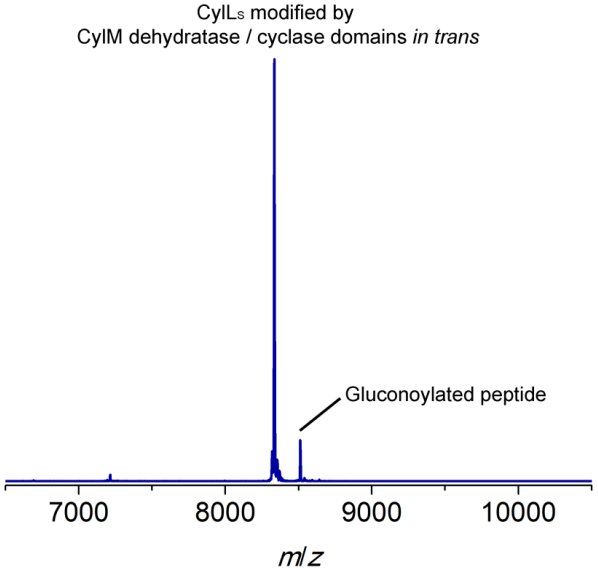

Figure 4.11 MALDI-TOF mass spectrum for $CylL_S$ modified by the CylM dehydratase domain and the CylM cyclase domain *in trans* in *E. coli*. Calculated $M–4H_2O$: 8,330, average mass; observed $M–4H_2O+H^+$: 8,334, average mass. Partial gluconoylation at the N terminus of $CylL_S$ occurred when expressing the peptide in *E. coli* BL21(DE3), resulting in a +178 Da peak in addition to the peak with the desired mass (*22*).

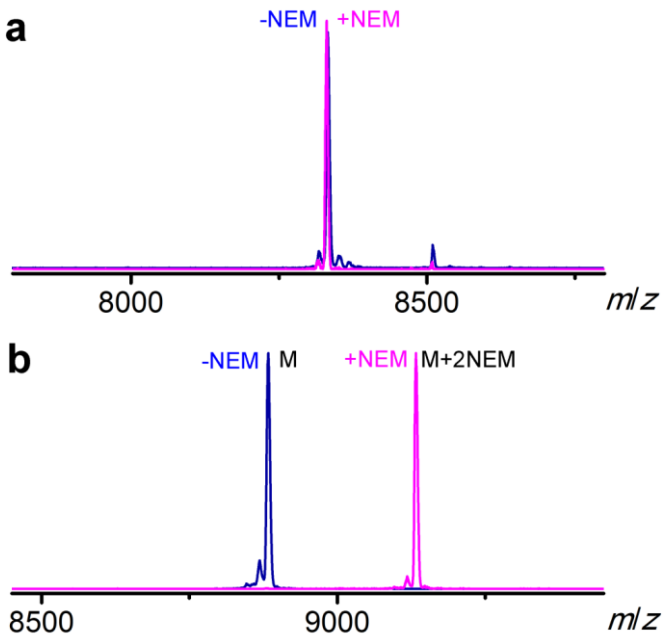

Figure 4.12 NEM analysis of $CylL_S$ modified by CylM dehydratase domain and CylM cyclase domain *in trans* in *E. coli*. Overlaid MALDI-TOF mass spectra of $CylL_S$ modified by CylM dehydratase domain and CylM cyclase domain *in trans* (**a**) (blue trace) and the peptide treated with NEM (magenta trace). Unmodified $CylL_S$ (**b**) was employed as a control. Two NEM adducts were observed for linear $CylL_S$ that contains two free thiols, whereas no NEM adducts were detected for $CylL_S$ modified by CylM dehydratase domain and CylM cyclase domain *in trans*, indicating the peptide was fully cyclized. For $CylL_S$ modified by CylM dehydratase domain and CylM cyclase domain *in trans*, calculated $M–4H_2O$: 8,330, average mass; observed $M–4H_2O+H^+$: 8,334, average mass; for such a peptide treated with NEM, calculated $M–4H_2O$+0NEM: 8,330, average mass; observed $M–4H_2O$+0NEM$+H^+$: 8,331, average mass. For unmodifed $CylL_S$, calculated M: 8,883, average mass; observed $M+H^+$: 8,883, average mass; for unmodifed $CylL_S$ treated by NEM, calculated M+2NEM: 9,133, average mass; observed M+2NEM$+H^+$: 9,133, average mass.

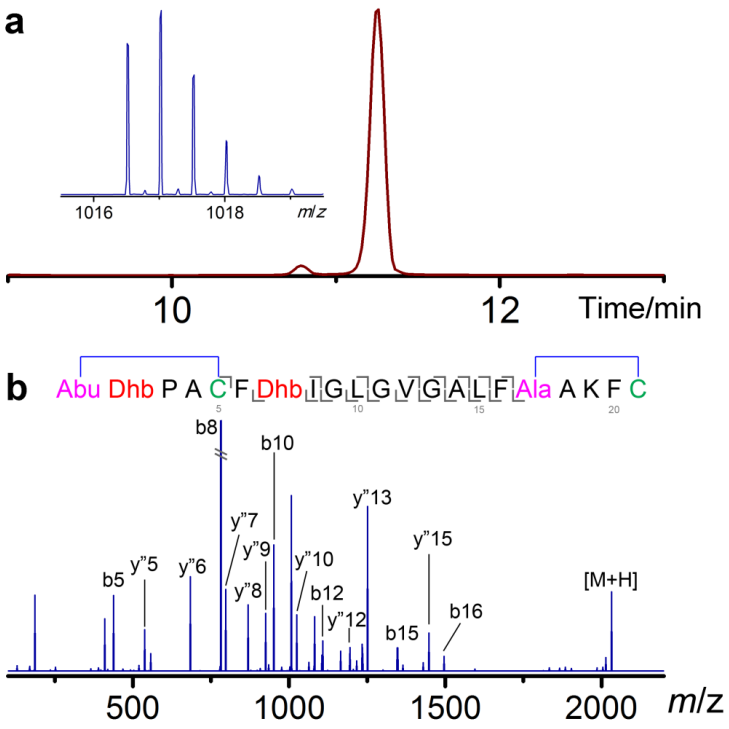

Figure 4.13 ESI MS analyses of cytolysin S peptide modified by CylM dehydratase domain and CylM cyclase domain *in trans* in *E. coli*. (**a**) Extracted ion chromatogram (*m*/*z* = 1017) of the LC trace for $CylL_S$ core peptide modified by CylM dehydratase domain and CylM cyclase domain *in trans*. The mass spectrum corresponding to the product peak is shown in the insert. For fully modified $CylL_S$ core peptide, calculated $[M–4H_2O+2H]^{2+}$: 1,016.52, monoisotopic mass; observed $[M–4H_2O+2H]^{2+}$: 1,016.51, monoisotopic mass. (**b**) MS/MS analysis of $CylL_S$ core peptide that was modified by CylM dehydratase domain and CylM cyclase domain *in trans* and a proposed structure deduced from the fragmentation pattern.

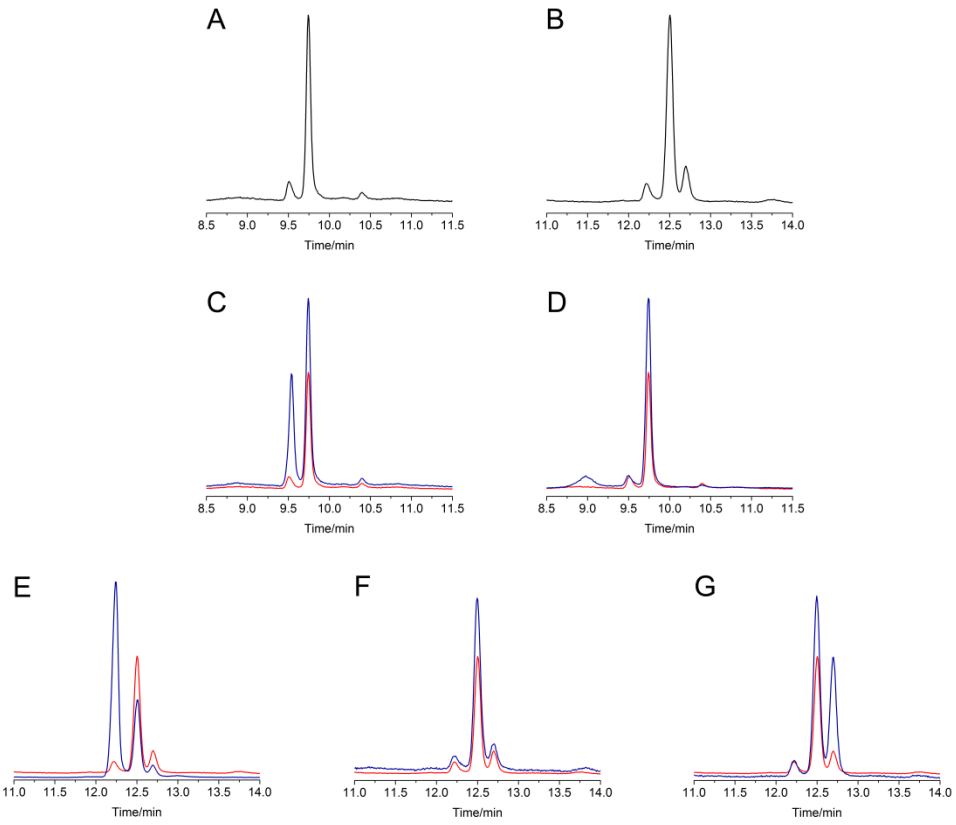

Figure 4.14 GC-MS traces for hydrolyzed/derivatized Lan/MeLan residues obtained from CylL$_S$ peptide modified by CylM dehydratase domain and CylM cyclase domain *in trans* in *E. coli* and co-injections with derivatized Lan/MeLan standards (selected ion monitoring, SIM, at 365 Da for Lan and 379 Da for MeLan). **(a)** Hydrolyzed and derivatized MeLan residues from the target peptide. **(b)** Hydrolyzed and derivatized Lan residues from the target peptide. **(c)** Hydrolyzed and derivatized MeLan residues from the target peptide (red line) and co-injected with derivatized DL-MeLan standard (blue line). **(d)** Hydrolyzed and derivatized MeLan residues from the target peptide (red line) and co-injected with derivatized LL-MeLan standard (blue line). **(e)** Hydrolyzed and derivatized Lan residues from the target peptide (red line) and co-injected with derivatized DD-Lan standard (blue line). **(f)** Hydrolyzed and derivatized Lan residues from the target peptide (red line) and co-injected with derivatized DL-Lan standard (blue line). **(g)** Hydrolyzed and derivatized Lan residues from the target peptide (red line) and co-injected with derivatized LL-Lan standard (blue line). Traces of derivatized Lan and MeLan residues from the target peptide (red lines) used for overlay were adjusted to 70% intensity for clarity. A small amount of epimerization of (Me)Lan occurs during acid hydrolysis of the peptides as has been reported previously (*23*).

As described in chapter 2, the formation of the unusual LL stereochemistry has been linked to a "Dhx-Dhx-Xxx-Xxx-Cys" motif present in the substrate peptide, and the second dehydroamino acid in such a motif was shown to be important for the stereoselective cyclization of the A ring of another lanthipeptide haloduracin β (Halβ) (*16*). To test whether the second Dhb is important for CylM-catalyzed cyclization of the "Dhb-Dhb-Pro-Ala-Cys" sequence (A ring of dehydrated $CylL_S$), Thr2 was substituted by Ala in $CylL_S$ to investigate if such a mutation would result in an inversion of the original LL stereochemistry. $CylL_S$-T2A peptide was modified by CylM in *E. coli*, and the resulting peptide was purified. MALDI-TOF MS analysis indicated the desired loss of three water molecules and no free thiols were present in the peptide (**Figures 4.15** and **4.16**). The modified $CylL_S$-T2A peptide was then hydrolyzed in acid, derivatized and analyzed by GC-MS. Surprisingly, signals corresponding to LL-MeLan and DL-Lan were exclusively detected (**Figure 4.17**), similar to what is observed for wild type cytolysin S. The selective formation of LL-MeLan in the A ring of $CylL_S$-T2A by CylM can be explained in two ways: 1) the second Dhb is not a crucial component for the "Dhb-Dhb-Pro-Ala-Cys" sequence to form an LL-MeLan; and/or 2) the synthetase CylM has a preference for the LL stereochemistry and catalyzes the formation of LL-MeLan even when the second Dhb is missing in the substrate motif.

To attempt to differentiate these two possibilities, I employed the non-enzymatic cyclization system described above and expressed $CylL_S$-T2A together with the CylM dehydratase domain in *E. coli*. As expected three dehydrations were observed, corresponding to fully dehydrated $CylL_S$-T2A (**Figure 4.18**). However, NEM analysis indicated one of the two thiols in the peptide remained free, different from what was observed previously for wild type $CylL_S$ modified by the CylM dehydratase domain and for $CylL_S$-T2A modified by full length CylM (**Figure 4.19**). Further removal of the leader peptide and MS analysis of the resulting mutated core peptide confirmed that it was the A ring that was not properly cyclized as hydrolysis of the N-terminal Dhb to 2-oxobutyrate was exclusively observed (**Figure 4.20**). The partially cyclized $CylL_S$-T2A peptide was purified and non-enzymatic cyclization of the A ring was pursued *in vitro*. Unfortunately, little ring closure activity was observed after 15 hours of incubation at pH 9 (**Figure 4.21**). Collectively, these results suggest that the cyclization of the A ring of dehydrated $CylL_S$-T2A requires the cyclase activity of CylM, and that non-enzymatic cyclization is greatly aided by the second Dhb in the "Dhb-Dhb-Pro-Ala-Cys" motif.

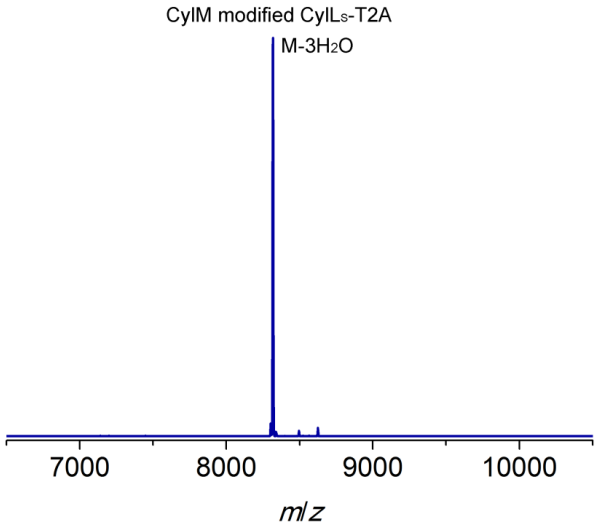

Figure 4.15 MALDI-TOF mass spectrum of CylL$_S$-T2A modified by CylM in *E. coli*. Calculated M–3H$_2$O: 8,318, average mass; observed M–3H$_2$O+H$^+$: 8,318, average mass.

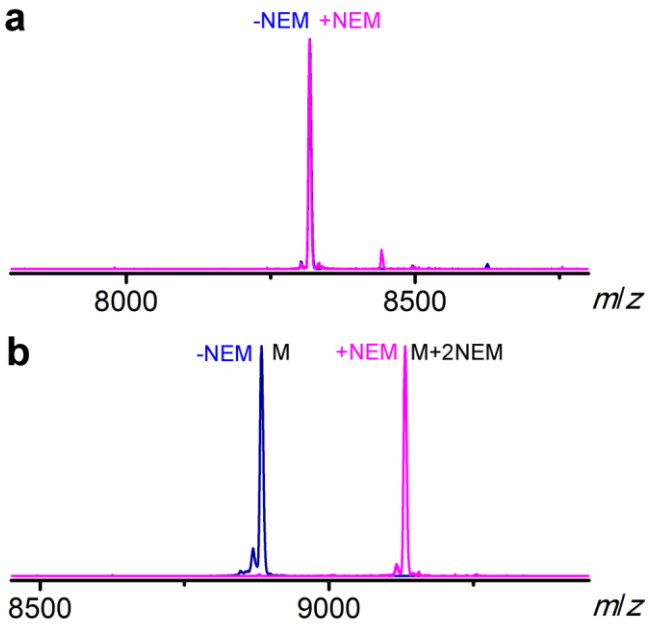

Figure 4.16 NEM analysis of CylM-modified CylL$_S$-T2A peptide. Overlaid MALDI-TOF mass spectra of CylM-modified CylL$_S$-T2A (**a**) (blue trace) and the peptide treated with NEM (magenta trace). Unmodified CylL$_S$ (**b**) was employed as a control. Two NEM adducts were observed for linear CylL$_S$ that contains two free thiols, whereas no NEM adducts were detected for CylM-modified CylL$_S$-T2A, indicating that both cysteines in the peptides are involved in (Me)Lan rings. For CylM-modified CylL$_S$-T2A, calculated M–3H$_2$O: 8,318, average mass; observed M–3H$_2$O+H$^+$: 8,319, average mass; for CylM-modified CylL$_S$-T2A treated by NEM, calculated M–3H$_2$O+0NEM: 8,318, average mass; observed M–3H$_2$O+0NEM+H$^+$: 8,317, average mass. For unmodifed CylL$_S$, calculated M: 8,883, average mass; observed M+H$^+$: 8,883, average mass; for unmodifed CylL$_S$ treated by NEM, calculated M+2NEM: 9,133, average mass; observed M+2NEM+H$^+$: 9,132, average mass.

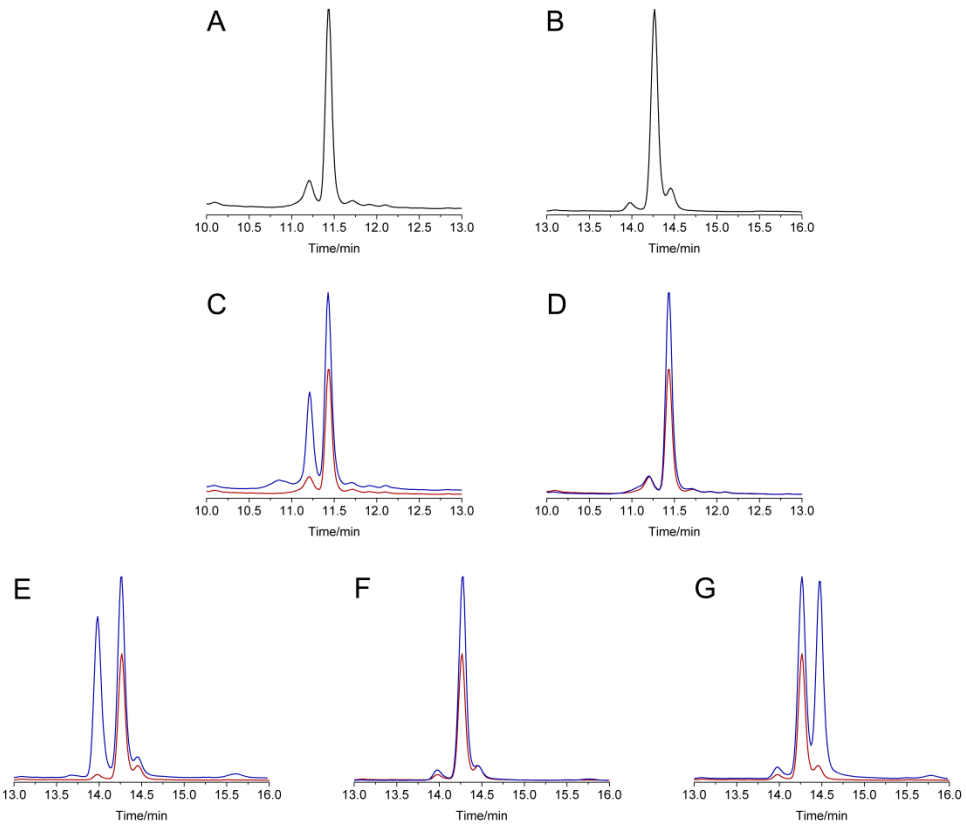

Figure 4.17 GC-MS traces for hydrolyzed and derivatized Lan/MeLan residues obtained from CylM-modified CylL$_S$–T2A in *E. coli* and co-injections with derivatized Lan/MeLan standards (selected ion monitoring, SIM, at 365 Da for Lan and 379 Da for MeLan). **(a)** Hydrolyzed and derivatized MeLan residues from CylM-modified CylL$_S$–T2A. **(b)** Hydrolyzed and derivatized Lan residues from CylM-modified CylL$_S$–T2A. **(c)** Hydrolyzed and derivatized MeLan residues from CylM-modified CylL$_S$–T2A (red line) and co-injected with derivatized DL-MeLan standard (blue line). **(d)** Hydrolyzed and derivatized MeLan residues from CylM-modified CylL$_S$–T2A (red line) and co-injected with derivatized LL-MeLan standard (blue line). **(e)** Hydrolyzed and derivatized Lan residues from CylM-modified CylL$_S$–T2A (red line) and co-injected with derivatized DD-Lan standard (blue line). **(f)** Hydrolyzed and derivatized Lan residues from CylM-modified CylL$_S$–T2A (red line) and co-injected with derivatized DL-Lan standard (blue line). **(g)** Hydrolyzed and derivatized Lan residues from CylM-modified CylL$_S$–T2A (red line) and co-injected with derivatized LL-Lan standard (blue line). Traces of derivatized Lan and MeLan residues from CylM-modified CylL$_S$–T2A (red lines) used for overlay were adjusted to 70% intensity for clarity. A small amount of epimerization of (Me)Lan occurs during acid hydrolysis of the peptides as has been reported previously (*23*).

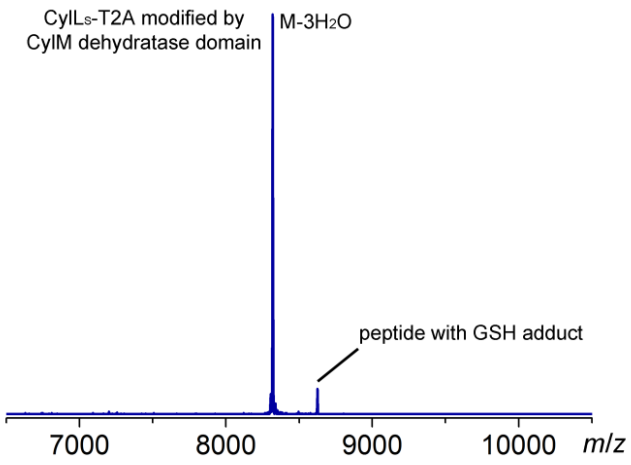

Figure 4.18 MALDI-TOF mass spectrum for $CylL_S$-T2A peptide modified by the CylM dehydratase domain in *E. coli*. Calculated M–$3H_2O$: 8,318, average mass; observed M–$3H_2O$+H^+: 8,319, average mass. A small +307 peak is attributed to a glutathione (GSH) addition to the peptide.

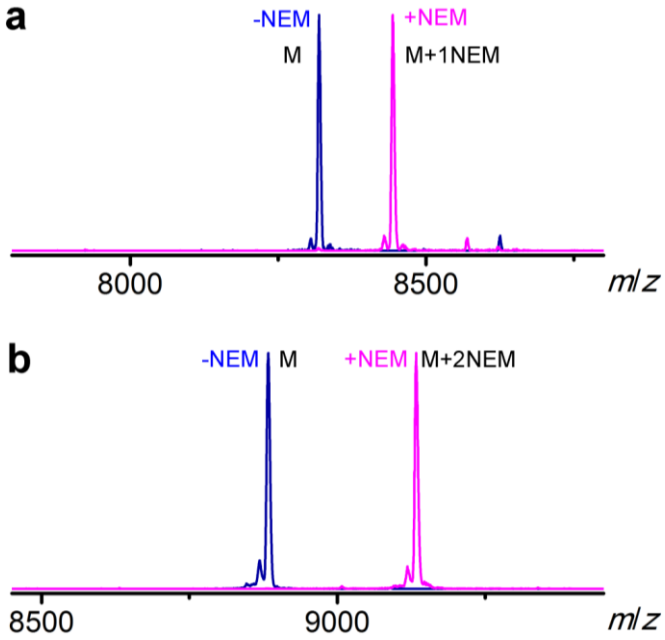

Figure 4.19 NEM analysis of $CylL_S$-T2A peptide modified by the CylM dehydratase domain in *E. coli*. Overlaid MALDI-TOF mass spectra of $CylL_S$-T2A modified by the CylM dehydratase domain (**a**) (blue trace) and the peptide treated with NEM (magenta trace). Unmodified $CylL_S$ (**b**) was employed as a control. Two NEM adducts were observed for linear $CylL_S$ that contains two free thiols, whereas one NEM adduct was detected for $CylL_S$-T2A modified by the CylM dehydratase domain, indicating that one of the two cysteines in the peptide was free. For $CylL_S$-T2A modified by the CylM dehydratase domain, calculated M–$3H_2O$: 8,318, average mass; observed M–$3H_2O$+H^+: 8,319, average mass; for $CylL_S$-T2A modified by CylM dehydratase domain and treated with NEM, calculated M–$3H_2O$+1NEM: 8,443, average mass; observed M–$3H_2O$+1NEM+H^+: 8,444, average mass. For unmodifed $CylL_S$, calculated M: 8,883, average mass; observed M+H^+: 8,883, average mass; for unmodifed $CylL_S$ treated by NEM, calculated M+2NEM: 9,133, average mass; observed M+2NEM+H^+: 9,133, average mass.

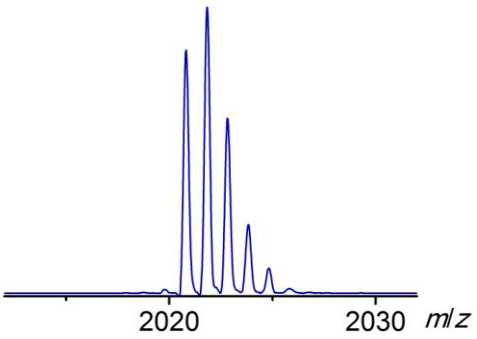

Figure 4.20 MALDI-TOF mass spectrum of the core peptide of $CylL_S$–T2A modified by the CylM dehydratase domain in *E. coli*. Hydrolysis of Dhb1 to 2-oxobutyrate resulted in a mass increase of 1 Da. Calculated M–$3H_2O$–NH+O: 2,020.0, monoisotopic mass; observed M–$3H_2O$–NH+O+H^+: 2,020.8, monoisotopic mass.

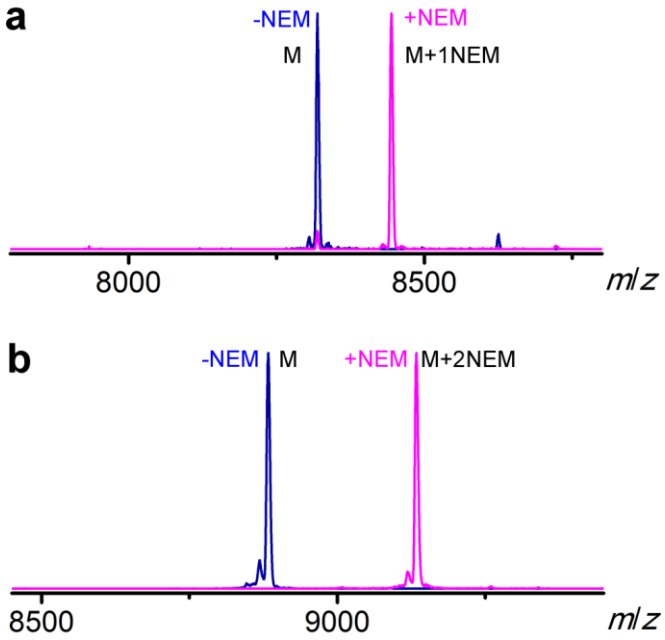

Figure 4.21 Non-enzymatic cyclization of dehydrated and partially cyclized $CylL_S$-T2A peptide *in vitro*. Overlaid MALDI-TOF mass spectra of $CylL_S$-T2A modified by CylM dehydratase domain (**a**) (blue trace) and the peptide incubated at basic pH to facilitate non-enzymatic cyclization and treated with NEM (magenta trace). Unmodified $CylL_S$ (**b**) was employed as a control. Two NEM adducts were observed for linear $CylL_S$ that contains two free thiols, whereas one NEM adduct was dominantly detected for the target peptide, indicating that non-enzymatic cyclization of the A ring hardly proceeded under the tested condition. For dehydrated and partially cyclized $CylL_S$-T2A, calculated M–$3H_2O$: 8,318, average mass; observed M–$3H_2O$+H^+: 8,319, average mass; for dehydrated and partially cyclized $CylL_S$-T2A treated with NEM, calculated M–$3H_2O$+1NEM: 8,443, average mass; observed M–$3H_2O$+1NEM+H^+: 8,444, average mass. For unmodifed $CylL_S$, calculated M: 8,883, average mass; observed M+H^+: 8,883, average mass; for unmodifed $CylL_S$ treated by NEM, calculated M+2NEM: 9,133, average mass; observed M+2NEM+H^+: 9,134, average mass.

These results seem to argue against to the first explanation as the property of the mutated A ring of $CylL_S$ is clearly different from the original sequence. However, it is not conclusive whether the capability of cyclizing the "Dhb-Ala-Pro-Ala-Cys" sequence in $CylL_S$-T2A to form an LL-MeLan residue is a specific property of CylM that differs from other LanM proteins. To answer this question, I tested whether a different LanM, the Halβ synthetase HalM2, would modify $CylL_S$-T2A to form an LL-MeLan in the A ring. A gene encoding a chimeric peptide with the leader peptide of Halβ connected to the $CylL_S$-T2A core peptide was constructed and coexpressed with *halM2* in *E. coli*. HalM2 carried out the desired three dehydrations of the $CylL_S$-T2A core peptide (**Figure 4.22**). The efficiency of the cyclization reaction was evaluated using the NEM assay. Partially cyclized HalA2-$CylL_S$-T2A peptide with one free thiol was detected almost exclusively (**Figure 4.23**), suggesting that HalM2 is inefficient in catalyzing the cyclization of the A ring of dehydrated $CylL_S$-T2A. Further GC-MS analysis indicated that only a very small amount of A ring was formed as much lower signal intensity was observed for MeLan compared to Lan. For the peptide with a cyclized A ring, both LL- and DL-MeLan were detected and the DL-MeLan-containing peptide appeared to be the major product (60%) (**Figure 4.24**). As expected, a signal corresponding to DL-Lan was exclusively detected for Lan on the GC chromatogram (**Figure 4.24**). These data suggest that HalM2 is inefficient in cyclizing the "Dhb-Ala-Pro-Ala-Cys" sequence in the $CylL_S$-T2A peptide, and that the stereoselectivity of HalM2 with respect to such a substrate sequence is different from that of CylM.

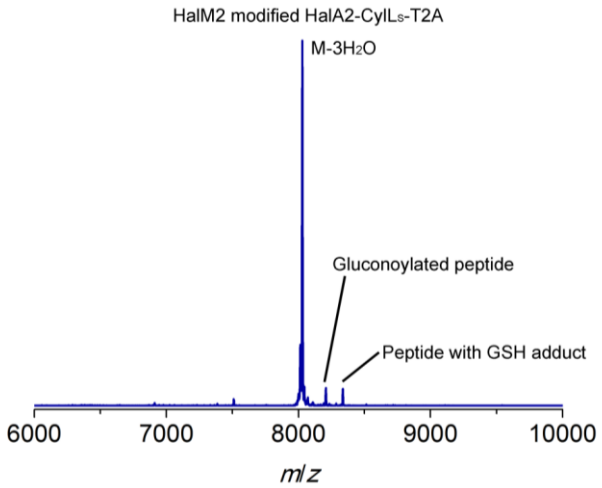

Figure 4.22 MALDI-TOF mass spectrum for HalA2-CylL$_S$-T2A modified by HalM2 in *E. coli*. Calculated M–3H$_2$O: 8,029, average mass; observed M–3H$_2$O+H$^+$: 8,029, average mass. Partial gluconoylation at the N terminus occurred when expressing the peptide in *E. coli* BL21(DE3), resulting in a +178 Da peak in addition to the peak with the desired mass (*22*). The small +307 peak is attributed to a glutathione (GSH) addition to HalM2-modified HalA2-CylL$_S$-T2A peptide.

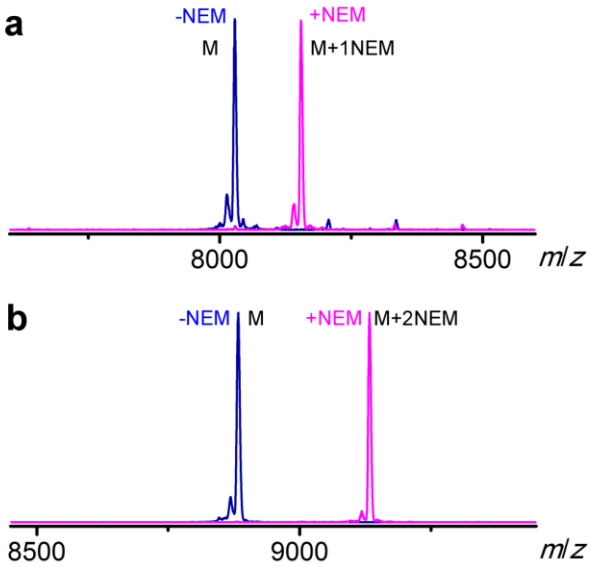

Figure 4.23 NEM analysis of HalA2-CylL$_S$-T2A peptide modified by HalM2 in *E. coli*. Overlaid MALDI-TOF mass spectra of HalM2-modified HalA2-CylL$_S$-T2A (**a**) (blue trace) and the peptide treated with NEM (magenta trace). Unmodified CylL$_S$ (**b**) was employed as a control. Two NEM adducts were observed for linear CylL$_S$ that contains two free thiols, whereas one NEM adduct was detected for HalM2-modified HalA2-CylL$_S$-T2A almost exclusively, indicating that one of the two cysteines in the peptide was not cyclized. For HalM2-modified CylL$_S$-T2A, calculated M–3H$_2$O: 8,029, average mass; observed M–3H$_2$O+H$^+$: 8,029, average mass; for HalM2-modified CylL$_S$-T2A treated with NEM, calculated M–3H$_2$O+1NEM: 8,154, average mass; observed M–3H$_2$O+1NEM+H$^+$: 8,155, average mass. For unmodifed CylL$_S$, calculated M: 8,883, average mass; observed M+H$^+$: 8,883, average mass; for unmodifed CylL$_S$ treated by NEM, calculated M+2NEM: 9,133, average mass; observed M+2NEM+H$^+$: 9,133, average mass.

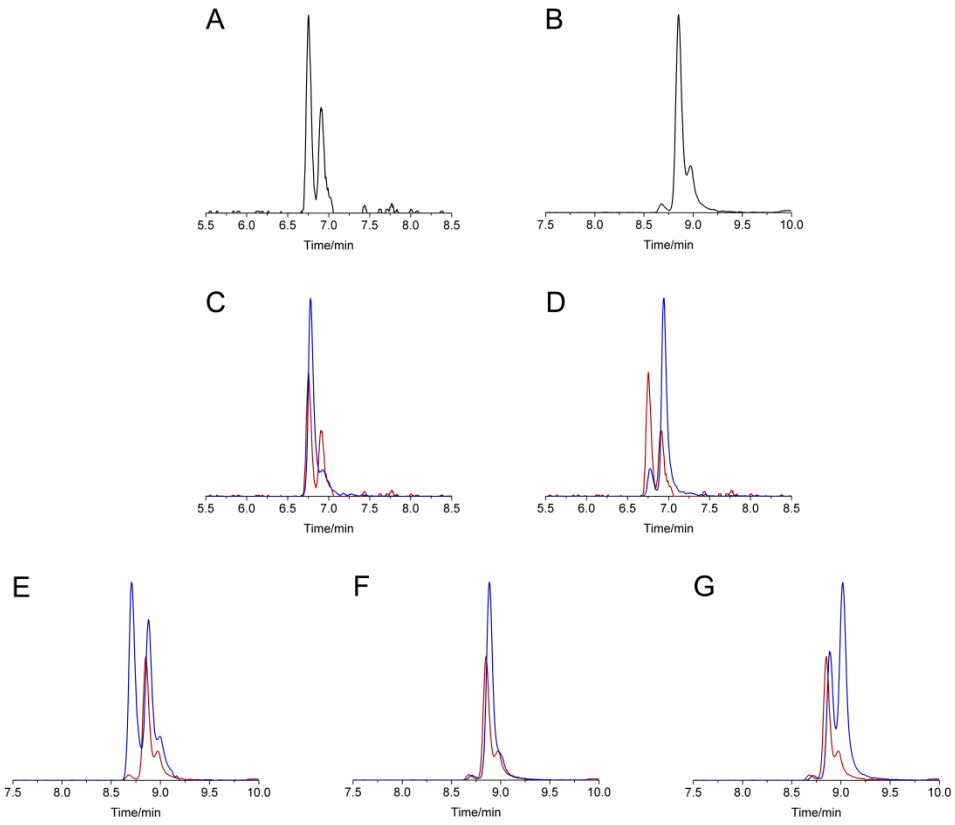

Figure 4.24 GC-MS traces for hydrolyzed/derivatized Lan/MeLan residues obtained from HalM2-modified HalA2-CylL$_S$–T2A in *E. coli* and co-injections with derivatized Lan/MeLan standards (selected ion monitoring, SIM, at 365 Da for Lan and 379 Da for MeLan). **(a)** Hydrolyzed and derivatized MeLan residues from HalM2-modified HalA2-CylL$_S$–T2A. **(b)** Hydrolyzed and derivatized Lan residues from HalM2-modified HalA2-CylL$_S$–T2A. **(c)** Hydrolyzed and derivatized MeLan residues from HalM2-modified HalA2-CylL$_S$–T2A (red line) and co-injected with derivatized DL-MeLan standard (blue line). **(d)** Hydrolyzed and derivatized MeLan residues from HalM2-modified HalA2-CylL$_S$–T2A (red line) and co-injected with derivatized LL-MeLan standard (blue line). **(e)** Hydrolyzed and derivatized Lan residues from HalM2-modified HalA2-CylL$_S$–T2A (red line) and co-injected with derivatized DD-Lan standard (blue line). **(f)** Hydrolyzed and derivatized Lan residues from HalM2-modified HalA2-CylL$_S$–T2A (red line) and co-injected with derivatized DL-Lan standard (blue line). **(g)** Hydrolyzed and derivatized Lan residues from HalM2-modified HalA2-CylL$_S$–T2A (red line) and co-injected with derivatized LL-Lan standard (blue line). Traces of derivatized Lan and MeLan residues from HalM2-modified HalA2-CylL$_S$–T2A (red lines) used for overlay were adjusted to 70% intensity for clarity. A small amount of epimerization of Lan occurs during acid hydrolysis of the peptides as has been reported previously (*23*).

4.3 Discussion

In this chapter, I reconstituted the non-enzymatic cyclization of $CylL_S$, both in *E. coli* and *in vitro* using a dehydrated peptide and confirmed that the dehydratase and cyclase activities of CylM can be independently accessed using truncated CylM proteins. The *in vitro* auto-cyclization of dehydrated $CylL_S$ proceeded efficiently at nearly neutral pH, which was unexpected before this work because of results obtained for another group of lanthipeptides, the prochlorosins, where the non-enzymatically catalyzed Michael-type addition required a higher pH to activate the thiols and was extremely slow, especially for the formation of MeLan residues (*24*). These observations suggest that dehydrated $CylL_S$ adopts a pre-organized conformation in which the cysteine thiols have increased reactivity towards the dehydroamino acids, which allows the cyclization to proceed even in the absence of a catalyst. Previous computational simulation of the dehydrated A ring of $CylL_S$ demonstrated that the overall geometry of the optimal transition state structure of such a sequence for a *Re*-face cyclization efficiently stabilizes the negative charge of the enolate intermediate (*16*), which accounts significantly for the lower activation energy barrier for the Michael-type addition to yield LL stereochemistry and might be the reason why this reaction can proceed efficiently without a cyclase.

Moreover, the formation of the correct ring topology for $CylL_S$ by non-enzymatic cyclization supports a previous hypothesis of substrate-controlled selectivity for lanthipeptide biosynthesis and suggests that both regioselectivity and stereoselectivity for cytolysin S synthesis are encoded in the $CylL_S$ sequence (*25*). There are in total 144 possible products for the cyclization process of dehydrated $CylL_S$, considering different ring topologies and stereochemistries of (Me)Lan residues with a fixed L configuration for the former Cys residue that does not alter during the process. However, only one major product was observed for $CylL_S$ that was cyclized non-enzymatically, indicating that dehydrated $CylL_S$ has a sufficiently well-organized scaffold that the preferred product out competes the other 143 possibilities and dominates the reaction (**Figures 4.25** and **4.26**). Collectively, these observations reinforce the idea of substrate-controlled reactivity and selectivity for lanthipeptide biosynthesis.

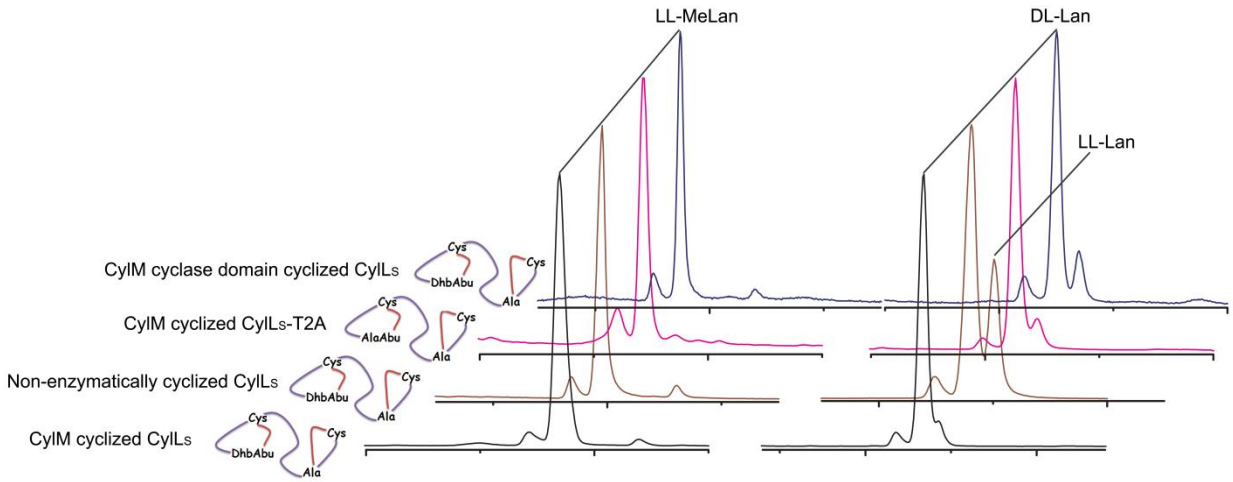

Figure 4.25 GC-MS traces for hydrolyzed and derivatized (Me)Lan residues from CylL$_S$ and CylL$_S$–T2A peptides that are cyclized using different set ups. Gas chromatograms showing signals corresponding to derivatized MeLan (left panel) and Lan (right panel) residues from CylL$_S$ peptides cyclized by CylM (black trace) (9), by the CylM cyclase domain (blue trace) or without a cyclase (brown trace), and CylL$_S$–T2A peptide cyclized by CylM (magenta trace) were overlaid.

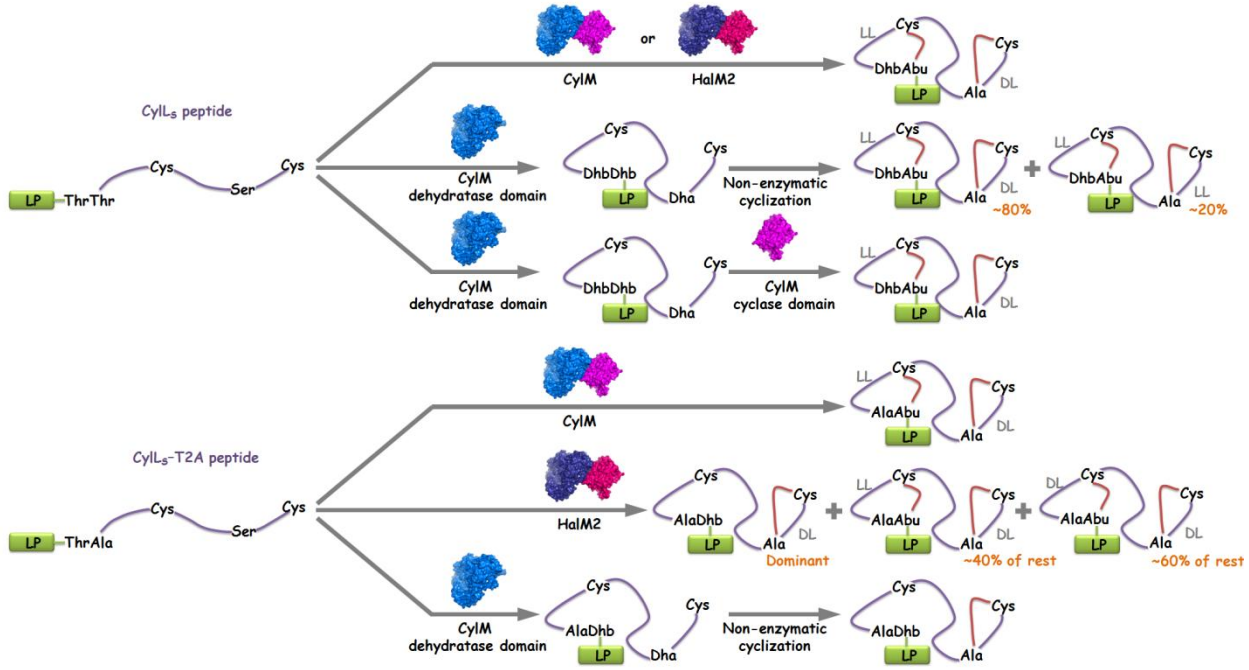

Figure 4.26 Summary of the CylL$_S$ and CylL$_S$–T2A peptides that are cyclized under different conditions. Generic protein structures are shown for CylM, HalM2, CylM dehydratase domain and CylM cyclase domain. Ring topologies of the final products were deduced form MS/MS fragmentation patterns, whereas configurations of (Me)Lan rings were assigned using GC-MS. Ratio of product mixture was estimated based on integration results of LC trace (for non-enzymatically cyclized CylL$_S$) or GC trace (for HalM2-modified CylL$_S$–T2A peptide with cyclized A ring). LP: leader peptide.

The second dehydroamino acid in a "Dhx-Dhx-Xxx-Xxx-Cys" motif has been proposed to be crucial for the formation of an LL-(Me)Lan as mutation of the second Dhb to Ala resulted in an inversion of the LL stereochemistry to DL for the A ring of Halβ cyclized by HalM2 (*9*). However, CylM cyclizes the A ring of $CylL_S$-T2A to form an LL-MeLan residue even when the second Dhb is missing (**Figures 4.25** and **4.26**). Unfortunately, the "Dhb-Ala-Pro-Ala-Cys" sequence was very unreactive towards non-enzymatic cyclization and cyclization by HalM2 (**Figure 4.26**). However, the limited amount of HalM2-cyclized product contained predominantly the DL stereochemistry. These observations imply coevolution of CylM with its substrate to promote stereoselective MeLan synthesis in the A ring of $CylL_S$. This coevolution phenomenon distinguishes CylM from other lanthionine synthetases, including those that also catalyze the formation of LL stereochemistry in their substrates. Specifically, HalM2 appears to not have been optimized because it could no longer maintain the LL configuration when the second dehydroamino acid is missing in HalA2. In this case, the synthetase may only be responsible for activating the cysteine thiol for the Michael-type addition, whereas the selective formation of the LL stereochemistry is only induced by the substrate sequence (**Figure 4.27**). A minor product corresponding to a diastereomer of cytolysin S with an LL B ring was observed for the non-enzymatically cyclized $CylL_S$, which was not detected if the CylM cyclase domain was supplied in the reaction (**Figures 4.25** and **4.26**), suggesting a role of the cyclase domain of CylM with respect to the stereoselective ring closure for the B ring of $CylL_S$. Collectively, these observations indicate that the stereoselectivity of CylM is more complicated than a simple preference for a certain configuration.

The reason why CylM coevolved with its substrates to introduce the LL stereochemistry remains unknown. Both units of cytolysin contain the unusual LL stereochemistry, and cytolysin is unique in the lanthipeptide family for its lytic activity against eukaryotic cells (*7*, *9*). The unusual LL stereochemistry may therefore be an essential feature that accounts for the virulence of cytolysin. However, whether there is a structure-activity correlation between the unusual LL stereochemistry and the toxin activity of cytolysin requires further investigation.

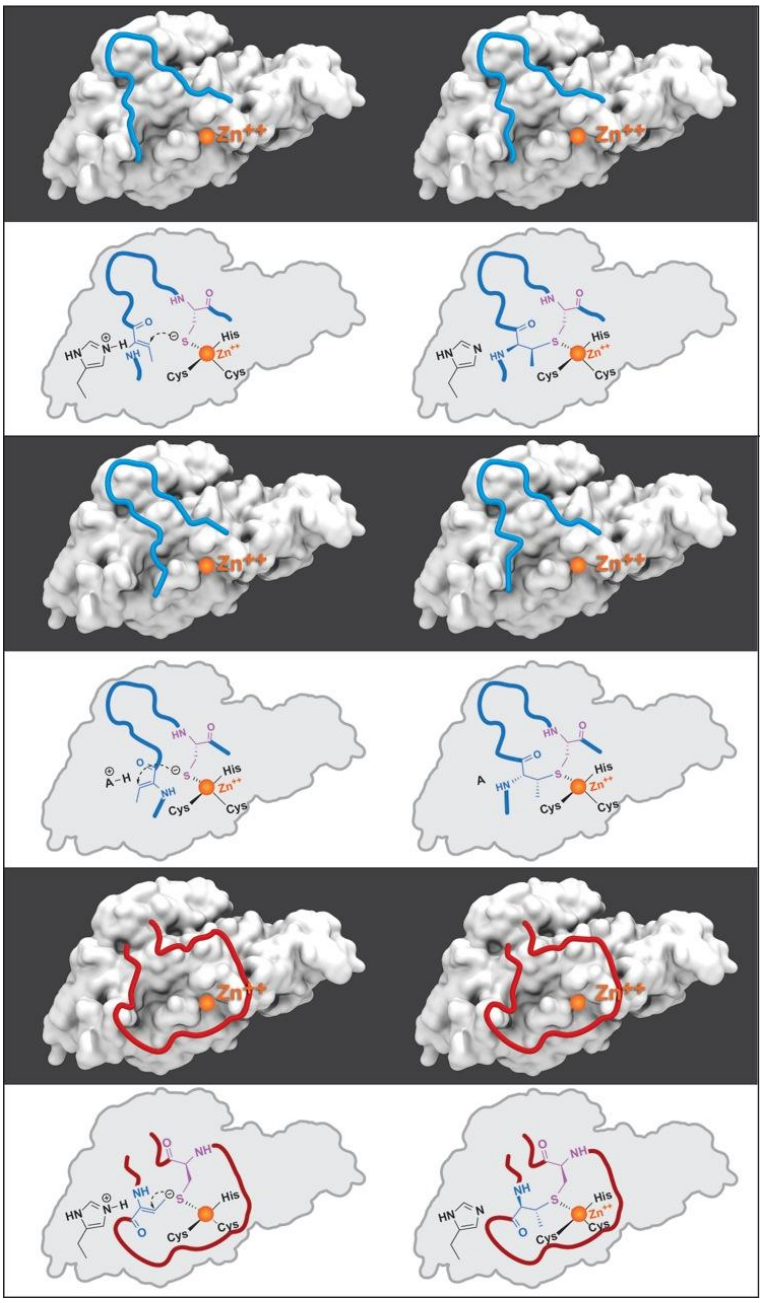

Figure 4.27 Hypothetic substrate patches on lanthionine synthetases that favor the formation of either DL or LL stereochemistry. In the proposed model, the majority of LanMs have substrate grooves on the protein surface that are optimized for the formation of DL-(Me)Lan (blue ribbons, upper panel). The specific motif "Dhx-Dhx-Xxx-Xxx-Cys" may introduce a turn in the substrate which forces a *Re* face addition onto the reacting Dhx, resulting in the formation of the LL stereochemistry (blue ribbons, middle panel). The stereoselectivity cannot be maintained once the structural-determining Dhx (non-reacting) is missing in the substrate as the enzyme groove has not been evolved to favor the formation of the LL stereochemistry. However, some LanMs, like CylM, may have coevolved with their substrates to contain a specific substrate patch optimized for LL ring closure (red ribbons, lower panel), which allows the formation of the unusual LL stereochemistry even the substrate induced turn is absent.

4.4 Methods

General methods

Similar general methods and materials as what were described in chapter 2 were employed in this chapter unless specified otherwise. LC-ESI-Q/TOF MS and MS/MS analyses were conducted using a Synapt G2S instrument with an Acquity UPLC system (Waters), which was equipped with a BEH C18 column (1.7 μm; 130Å; 50 x 2.1 mm; Waters). GC-MS analysis was performed in the Mass Spectrometry Laboratory (School of Chemical Sciences, UIUC) or the Roy J. Carver Metabolomics Center (UIUC) on an Agilent 7890 gas chromatograph with a CP-Chirasil-L-Val fused silica column (25 m x 0.25 mm x 0.15 μm) (Agilent). An Agilent 1260 Infinity Quaternary LC System was employed for analytical scale HPLC, whereas preparative HPLC was performed using a Waters Delta 600 instrument. Solid phase extraction was carried out with Strata-X polymeric RP columns (Phenomenex). Expression vector pET15b was obtained from Novagen.

Construction of the pRSFDuet-1/CylL$_S$/CylM-1-625-2 plasmid

The *cylL$_S$* gene was cloned into the multiple cloning site 1 (MCS1) of a pRSFDuet-1 vector, resulting in a pRSFDuet-1/CylL$_S$ plasmid. The *cylM-1-625* gene was amplified from the pRSFDuet-1/CylM plasmid and cloned into the MCS2 of pRSFDuet-1/CylL$_S$ to generate pRSFDuet-1/CylL$_S$/CylM-1-625-2. Primer sequences are listed in **Table 4.1**.

Construction of the pACYCDuet-1/ CylM-626-993-2 plasmid

The *cylM-626-993* gene was amplified from the pRSFDuet-1/CylM plasmid and cloned into the MCS2 of a pACYCDuet-1 vector to generate pACYCDuet-1/CylM-626-993-2. Primer sequences are listed in **Table 4.1**.

Construction of pRSFDuet-1 derivatives for coexpression of CylM with CylL$_S$-T2A, HalM2 with HalA2-CylL$_S$-T2A, and CylM-1-625 with CylL$_S$-T2A

The genes *cylL$_S$-T2A* and *halA2-CylL$_S$-T2A* are generated using overlap extension PCR based on *cylL$_S$* and *halA2* and cloned into the MCS1 of plasmids pRSFDuet-1/CylM-2 (*9*),

pRSFDuet-1/HalM2-2 (*26*) and pRSFDuet-1/CylM-1-625-2, respectively, resulting in pRSFDuet-1/CylL$_S$-T2A/CylM-2, pRSFDuet-1/HalA2-CylL$_S$-T2A/HalM2-2 and pRSFDuet-1/CylL$_S$-T2A/CylM-1-625-2. Primer sequences are listed in **Table 4.1**.

Table 4.1 Primer sequences used in this chapter.

Primer Name	Primer Sequence (5'-3')
HalA2_BamHI_FP	AAAAA GGATCC G ATGGTAAATT CAAAAGATTT
CylL$_S$_EcoRI_FP	AAAAA GAATTCG CTGAATA AAGAAAATCA G
CylL$_S$_NotI_RP	AAAAA GCGGCCGC TTAGCAA AATTTTGCGC T
CylL$_S$_BamHI_FP	AAAAA GGATCCG CTGAATA AAGAAAATCA G
CylL$_S$_HindIII_RP	AAAAA AAGCTT TTAGCAA AATTTTGCGC T
CylL$_S$_Leader_RP	T TCTGCCTGAA CATCACCGCT
CylL$_S$_T2A_FP	GATGTTCAGGCAGAAACCGCACCGGCATGTTTTACCATTGGT
HalA2_Leader_RP	TTGTGCAT GCACATCTCC TGA
HalA2_CylL$_S$_T2A _FP	GATGTGCATGCACAA ACC GCA CCGGCATGTTTTACC ATTGGT
CylM_NdeI_FP	AAAAA CATATG GAAGATA ATCTGATTAA T
CylM_KpnI_RP	AAAAA GGTACC TTACAGT TCAAACAGCA G
CylM625_KpnI_RP	AAAAA GGTACC TTA GTACGGGTTA TAAATATTCA G
CylM626_NdeI_FP	AAAAA CATATG TATATCAATGATCTGAAAAAC

Production and purification of His$_6$-CylL$_S$ modified by CylM dehydratase domain, His$_6$-CylL$_S$–T2A modified by CylM, His$_6$-CylL$_S$–T2A modified by CylM dehydratase domain and His$_6$-HalA2-CylL$_S$–T2A modified by HalM2

E. coli BL21 (DE3) cells were transformed with the pRSFDuet-1/CylL$_S$/CylM-1-625-2, pRSFDuet-1/CylL$_S$-T2A/CylM-2, pRSFDuet-1/HalA2-CylL$_S$-T2A/HalM2-2 and pRSFDuet-1/CylL$_S$-T2A/CylM-1-625-2 plasmids, and plated on an LB plate containing 50 mg/L kanamycin. A single colony was picked and grown in 10 mL of LB with 50 mg/L kanamycin aerobically at 37 ^oC for 10-16 h and the resulting culture was used to inoculate 1 L of LB. Cells were cultured at 37 ^oC until the OD at 600 nm reached 0.5, cooled to ~ 10 ^oC and IPTG was added to a final concentration of 0.1 mM. The culture was allowed to grow at 18 ^oC for another 18 h before harvesting. The cell pellet was resuspended at room temperature in LanA start buffer (500 mM NaCl, 0.5 mM imidazole, 20% glycerol, 20 mM NaH_2PO_4, pH 7.5 at 25 ^oC) or LanA buffer 1 (6 M guanidine hydrochloride, 500 mM NaCl, 0.5 mM imidazole, 20 mM NaH_2PO_4, pH 7.5 at 25 ^oC) followed by lysis via sonication. The lysed sample was then centrifuged at 23,700×*g* for 30 min and the supernatant was passed through 0.45-μm syringe filters. His-tagged

modified peptides were purified by immobilized metal affinity chromatography (IMAC) loaded with nickel as previously described (*27*). The eluted fractions with the desired peptide were desalted by preparative RP-HPLC using a Waters Delta-pak C4 column (15 μm; 300 Å; 25 x 100 mm) or by Strata-X polymeric RP solid phase extraction columns (Phenomenex). The desalted peptides were lyophilized and stored at −20 ^{o}C for future use.

Production and purification of His_6-$CylL_S$ modified by CylM dehydratase domain and CylM cyclase domain *in trans*

E. coli BL21 (DE3) cells were transformed with both pRSFDuet-1/$CylL_S$/CylM-1-625-2 and pACYCDuet-1/CylM-626-993-2 plasmids, and plated on an LB plate containing 50 mg/L kanamycin and 25 mg/L chloramphenicol. A single colony was picked for overexpression purposes. A similar procedure as that was described above was employed for expression and purification of the modified peptide.

Overexpression and purification of His_6-$CylL_S$ and His_6-CylM

Same procedure was followed for expression and purification of His_6-$CylL_S$ peptide and His_6-CylM protein as described in chapter 5 (*19*).

Production and purification of oxidized and dehydrated $CylL_S$

His_6-$CylL_S$ was expressed in *E. coli* and purified in its oxidized form with an intramolecular disulfide (see main text for more information). Due to high hydrophobicity, the peptide solubility in aqueous phase is poor. It could not dissolve completely in water even at low μM concentrations. As a result, a suspension of 2 mg/mL peptide was made for His_6-$CylL_S$ as a stock solution. The peptide stock was vortexed to a homogenized suspension each time before any peptide was taken. However, given the presence of precipitation, the peptide concentration could not be tightly controlled.

To a reaction vessel containing 50 mM HEPES (pH = 7.5), 10 mM $MgCl_2$, 5 mM ATP and 10^{-5} U thrombin (for in situ removal of the His-tag), 150 μM of His_6-$CylL_S$ was supplied followed by the addition of CylM to a final concentration of 2 μM. The reaction was allowed to proceed at room temperature for more than 24 h and complete loss of four water molecules in

$CylL_S$ was confirmed by MALDI-TOF MS. Subsequent separation of oxidized and dehydrated $CylL_S$ from CylM was achieved by analytical scale HPLC. Solvents used for the HPLC purification were: solvent A = 0.1% trifluoroacetic acid (TFA) in ddH_2O and solvent B = 0.086% TFA in 80% acetonitrile/20% ddH_2O. The reaction mixture was applied onto a Jupiter Proteo C12 column (5 μm; 90 Å; 250 x 4.6 mm; Phenomenex) and subsequent elution was carried out using a linear gradient from 2% to 100% B over 45 min. Fractions containing the desired peptide product were collected, lyophilized to dryness and kept at −20 ^oC until future usage.

In vitro non-enzymatic cyclization of dehydrated $CylL_S$

Oxidized and dehydrated $CylL_S$ obtained using the above described procedure was dissolved in ddH_2O to make a solution. The peptide could not be quantified by weighing as very limited amount of materials was obtained after HPLC purification. A concentration of 20~200 μg/mL was estimated based on the peak absorbance at 220 nm during HPLC purification. To such a peptide solution, 50 mM HEPES (pH = 7.5) and 5 mM DTT were added. The reaction was allowed to proceed at room temperature for 12 h before analysis. A parallel reaction with 20 μM unmodified $CylL_S$ was set up as a control.

For non-enzymatic cyclization of the A ring of dehydrated and partially cyclized $CylL_S$-T2A, 40 μM peptide was supplied to a reaction system containing 10 mM DTT and 50 mM Tris buffer (pH = 8.9). The reaction was incubated at room temperature for 15 h before analysis by NEM assay. A parallel reaction with 20 μM unmodified $CylL_S$ was set up as a control.

NEM analysis

For pure peptide samples, a two-step reaction was employed for NEM analysis, including a reduction step and a subsequent alkylation reaction. The typical set up for the reduction reaction included 0.3~0.6 mg/mL (30~80 μM) target peptides, 50 mM phosphate buffer (pH = 5.5) and 5 mM TCEP (tris(2-carboxyethyl)phosphine). The reaction mixture was kept at room temperature for 2 to 6 h to reduce all disulfide linkages in the peptide samples. Subsequent addition of 100 mM phosphate buffer (pH = 7.2) adjusted the pH of the reaction mixture to 6.5~7.0, to which NEM was added to a final concentration of 10 mM. The alkylation reaction was allowed to

proceed at 37 ^{o}C for 5 min before analysis by MALDI-TOF MS. Parallel reactions using unmodified His_6-$CylL_S$ (20 μM) were always performed as a positive control.

For NEM analysis of *in vitro* cyclized $CylL_S$, 10 mM NEM was directly added into a reaction aliquot (the peptide was in its reduced form) and incubated at 37 ^{o}C for 5 min before analysis. A parallel reaction using unmodified $CylL_S$ treated with the same procedure (described in the previous section) was employed as a positive control.

For NEM analysis of *in vitro* cyclization trials for $CylL_S$-T2A (for the A ring), 100 mM HEPES buffer (pH = 6.5) was added into a reaction aliquot to bring the pH to ~ 7, followed by the addition of 10 mM NEM. The alkylation reaction was incubated at 37 ^{o}C for 5 min before analysis. A parallel reaction using unmodified $CylL_S$ treated with the same procedure (described in the previous section) was employed as a positive control.

Leader peptide removal of $CylL_S$ and $CylL_S$-T2A peptides

To a peptide solution or a reaction mixture with a buffered pH around 7.5, 0.01~0.1 mg/mL CylA was added. DTT was supplied to a final concentration of 5 mM when desired. The reaction was allowed to proceed at room temperature for 1~6 h before analysis. Complete consumption of the full length peptide was confirmed by MALDI-TOF MS.

LC-ESI-Q/TOF MS and MS/MS analyses

Peptide samples of 100~500 ng (approximated by precursor peptide weight) were applied onto a BEH C18 column (1.7 μm; 130Å; 50 x 2.1 mm; Waters) that was pre-equilibrated in aqueous solvent A. The solvents used for LC were: solvent A = 0.1% formic acid in 95% ddH_2O/5% acetonitrile and solvent B = 0.1% formic in 5% ddH_2O/95% acetonitrile. Flow rate was set to 0,5 or 0.3 mL/min. A linear gradient from 0% to 60% B was employed from 0.5 min to 12.5 min and the fractionated sample was directly subjected to ESI-Q/TOF MS analysis using a Waters Synapt G2S mass spectrometer. The mass spectrometer was calibrated before any sample was injected. Data were acquired in ESI positive mode. The trap collision energy was set to 6 V for MS analysis. For MSMS analysis, a trap collision energy ramp ranging from 20-30 V was applied on 2+ parent ions to achieve fragmentation. Leucine-enkephalin was directly infused

as lock mass with lock spray sampling. The acquired spectra were processed using MaxEnt3 software and analyzed by Protein/Peptide Editor in BioLynx 4.1 (Waters).

GC-MS analysis

The synthesis of Lan and MeLan standards and the preparation of samples for GC-MS analysis were carried out following a reported procedure published elsewhere (*28, 29*). The modified full length peptides with their leader peptides attached were hydrolyzed and the resulting solutions were dried and directly used for derivatization. The derivatized samples were analyzed by GC-MS using an Agilent 7890 gas chromatograph equipped with a CP-Chirasil-L-Val fused silica column (25 m x 0.25 mm x 0.15 μm). Samples were dissolved in methanol and introduced to the instrument via a split (1:5 or 1:20) injection at a flow rate of 1.7 mL/min or 2.0 mL/min helium gas. The temperature method used was 160 ^{o}C for 2 min, 160 ^{o}C to 180 ^{o}C at 3 ^{o}C/min, and 180 ^{o}C for 2 min. The selected-ion monitoring (SIM) was achieved by extracting the total ion spectra for peaks containing the characteristic fragment masses of 365 Da for Lan and 379 Da for MeLan residues.

4.5 References

1. McIntosh, J. A., Donia, M. S., and Schmidt, E. W. (2009) Ribosomal peptide natural products: bridging the ribosomal and nonribosomal worlds, *Nat. Prod. Rep. 26*, 537–559.
2. Arnison, P. G., Bibb, M. J., Bierbaum, G., Bowers, A. A., Bugni, T. S., Bulaj, G., Camarero, J. A., Campopiano, D. J., Challis, G. L., Clardy, J., Cotter, P. D., Craik, D. J., Dawson, M., Dittmann, E., Donadio, S., Dorrestein, P. C., Entian, K.-D., Fischbach, M. A., Garavelli, J. S., Göransson, U., Gruber, C. W., Haft, D. H., Hemscheidt, T. K., Hertweck, C., Hill, C., Horswill, A. R., Jaspars, M., Kelly, W. L., Klinman, J. P., Kuipers, O. P., Link, A. J., Liu, W., Marahiel, M. A., Mitchell, D. A., Moll, G. N., Moore, B. S., Müller, R., Nair, S. K., Nes, I. F., Norris, G. E., Olivera, B. M., Onaka, H., Patchett, M. L., Piel, J., Reaney, M. J. T., Rebuffat, S., Ross, R. P., Sahl, H.-G., Schmidt, E. W., Selsted, M. E., Severinov, K., Shen, B., Sivonen, K., Smith, L., Stein, T., Süssmuth, R. E., Tagg, J. R., Tang, G.-L., Truman, A. W., Vederas, J. C., Walsh, C. T., Walton, J. D., Wenzel, S. C., Willey, J. M., and van der Donk, W. A. (2013) Ribosomally Synthesized and Post-translationally Modified Peptide Natural Products: Overview and Recommendations for a Universal Nomenclature, *Nat. Prod. Rep. 30*, 108-160.
3. Freeman, M. F., Gurgui, C., Helf, M. J., Morinaka, B. I., Uria, A. R., Oldham, N. J., Sahl, H. G., Matsunaga, S., and Piel, J. (2012) Metagenome mining reveals polytheonamides as posttranslationally modified ribosomal peptides, *Science 338*, 387-390.
4. Sit, C. S., McKay, R. T., Hill, C., Ross, R. P., and Vederas, J. C. (2011) The 3D structure of thuricin CD, a two-component bacteriocin with cysteine sulfur to alpha-carbon cross-links, *J. Am. Chem. Soc. 133*, 7680-7683.
5. Lohans, C. T., and Vederas, J. C. (2014) Structural characterization of thioether-bridged bacteriocins, *J. Antibiot. 67*, 23-30.
6. Knerr, P. J., and van der Donk, W. A. (2012) Discovery, biosynthesis, and engineering of lantipeptides, *Annu. Rev. Biochem. 81*, 479-505.
7. Bierbaum, G., and Sahl, H. G. (2009) Lantibiotics: mode of action, biosynthesis and bioengineering, *Curr. Pharm. Biotechnol. 10*, 2-18.
8. Chatterjee, C., Paul, M., Xie, L., and van der Donk, W. A. (2005) Biosynthesis and mode of action of lantibiotics, *Chem. Rev. 105*, 633-684.
9. Tang, W., and van der Donk, W. A. (2013) The sequence of the enterococcal cytolysin imparts unusual lanthionine stereochemistry, *Nat. Chem. Biol. 9*, 157-159.
10. Cox, C. R., Coburn, P. S., and Gilmore, M. S. (2005) Enterococcal cytolysin: A novel two component peptide system that serves as a bacterial defense against eukaryotic and prokaryotic cells, *Curr. Protein Pept. Sci. 6*, 77-84.
11. Todd, E. W. (1934) A comparative serological study of streptolysins derived from human and from animal infections, with notes on pneumococcal h ae molysin, tetanolysin and staphylococcus toxin, *J. Pathol. Bateriol. 39*, 299-321.

12. Ike, Y., and Clewell, D. B. (1984) Genetic analysis of the pAD1 pheromone response in *Streptococcus faecalis*, using transposon Tn917 as an insertional mutagen, *J. Bacteriol. 158*, 777-783.
13. Huycke, M. M., Spiegel, C. A., and Gilmore, M. S. (1991) Bacteremia caused by hemolytic, high-level gentamicin-resistant *Enterococcus faecalis*, *Antimicrob. Agents Chemother. 35*, 1626-1634.
14. Chow, J. W., Thal, L. A., Perri, M. B., Vazquez, J. A., Donabedian, S. M., Clewell, D. B., and Zervos, M. J. (1993) Plasmid-associated hemolysin and aggregation substance production contribute to virulence in experimental enterococcal endocarditis, *Antimicrob. Agents Chemother. 37*, 2474-2477.
15. Van Tyne, D., Martin, M. J., and Gilmore, M. S. (2013) Structure, function, and biology of the enterococcus faecalis cytolysin, *Toxins 5*, 895-911.
16. Tang, W., Jiménez-Osés, G., Houk, K. N., and van der Donk, W. A. (2015) Substrate control in stereoselective lanthionine biosynthesis, *Nat. Chem. 7*, 57-64.
17. Xie, L., Miller, L. M., Chatterjee, C., Averin, O., Kelleher, N. L., and van der Donk, W. A. (2004) Lacticin 481: in vitro reconstitution of lantibiotic synthetase activity, *Science 303*, 679-681.
18. You, Y. O., and van der Donk, W. A. (2007) Mechanistic investigations of the dehydration reaction of lacticin 481 synthetase using site-directed mutagenesis, *Biochemistry 46*, 5991-6000.
19. Dong, S.-H., Tang, W., Lukk, T., Nair, S. K., and van der Donk, W. A. (2015) The enterrococcal cytolysin synthetase has an unanticipated lipid kinase fold, *submitted for publication*.
20. Wagoner, J., Yu, Y., and Van der Donk, W. A. (2015) The leader peptide is not involved in the directionality of the lacticin 481 synthetase, *in preparation*.
21. Thibodeaux, C. J., Ha, T., and van der Donk, W. A. (2014) A price to pay for relaxed substrate specificity: a comparative kinetic analysis of the class II lanthipeptide synthetases ProcM and HalM2, *J. Am. Chem. Soc. 136*, 17513-17529.
22. Aon, J. C., Caimi, R. J., Taylor, A. H., Lu, Q., Oluboyede, F., Dally, J., Kessler, M. D., Kerrigan, J. J., Lewis, T. S., Wysocki, L. A., and Patel, P. S. (2008) Suppressing posttranslational gluconoylation of heterologous proteins by metabolic engineering of Escherichia coli, *Appl. Environ. Microbiol. 74*, 950-958.
23. Kido, Y., Hamakado, T., Yoshida, T., Anno, M., Motoki, Y., Wakamiya, T., and Shiba, T. (1983) Isolation and characterization of ancovenin, a new inhibitor of angiotensin I converting enzyme, produced by actinomycetes, *J. Antibiot. 36*, 1295-1299.
24. Mukherjee, S., and van der Donk, W. A. (2014) Mechanistic Studies on the Substrate-Tolerant Lanthipeptide Synthetase ProcM, *J. Am. Chem. Soc. 136*, 10450-10459.
25. Zhang, Q., Yu, Y., Velásquez, J. E., and van der Donk, W. A. (2012) Evolution of lanthipeptide synthetases, *Proc. Natl. Acad. Sci. U. S. A. 109*, 18361-18366.

26. Shi, Y., Yang, X., Garg, N., and van der Donk, W. A. (2011) Production of lantipeptides in *Escherichia coli*, *J. Am. Chem. Soc. 133*, 2338-2341.
27. Li, B., Cooper, L. E., and van der Donk, W. A. (2009) *In vitro* studies of lantibiotic biosynthesis *Methods Enzymol. 458*, 533-558.
28. Liu, W., Chan, A. S., Liu, H., Cochrane, S. A., and Vederas, J. C. (2011) Solid supported chemical syntheses of both components of the lantibiotic lacticin 3147, *J. Am. Chem. Soc. 133*, 14216-14219.
29. Tang, W., and van der Donk, W. A. (2012) Structural characterization of four prochlorosins: a novel class of lantipeptides produced by planktonic marine cyanobacteria, *Biochemistry 51*, 4271-4279.

Chapter 5. Mechanistic Studies of the Enterococcal Cytolysin Synthetase CylM

5.1 Introduction

Cytolysin is produced by many clinical isolates of *Enterococcus faecalis* and consists of two lanthipeptides (cytolysin L and S) (*1, 2*). For more details about the clinical importance of the enterococcal cytolysin and the biosynthesis of lanthipeptides, see chapters 1 and 2. The two-step post-translational modification of cytolysin that includes dehydration and cyclization of the precursor peptides $CylL_L$ and $CylL_S$ is catalyzed by CylM, a bifunctional lanthionine synthetase.

At present at least four distinct routes to lanthipeptides have been discovered, illustrating that the cyclic thioether motif is a privileged structural scaffold that has been independently accessed multiple times during evolution (*3*). These four routes differ primarily in the mechanism of dehydration (section 1.3 of chapter 1). For the class I, III, and IV lanthipeptides, the mechanism of dehydration has been illuminated by crystallographic characterization of the dehydratases or close sequence homologs (*4-7*), but the mechanism of dehydration for class II lanthipeptide synthetases that include CylM has remained enigmatic as the enzymes do not show any sequence homology with other proteins and structural information has not been available. In this work, the structure of CylM was determined, which demonstrates that the dehydration domain of this enzyme has structural similarity with eukaryotic lipid kinases despite the absence of notable sequence homology (crystalization and structural elucidation of CylM were performed by Dr. Shi-Hui Dong and Prof. Satish K. Nair). Structure-based generation of mutants provided insights into the mechanism of dehydration.

5.2 Results

5.2.1 *In vitro* reconstitution of CylM activity

Previously, I showed that CylM could dehydrate and cyclize its substrates, $CylL_L$ and $CylL_S$, in *Escherichia coli* (chapter 2) (*8*). To investigate the elusive mechanism of dehydration, I also reconstituted the CylM activity *in vitro*. CylM and its substrate peptides were expressed in *E. coli* as N-terminal hexahistidine-tagged proteins and purified by metal affinity chromatography (**Figure 5.1**). Incubation of His_6-CylM with $CylL_L$ or $CylL_S$ in the presence of dithiothreitol,

$MgCl_2$, and ATP demonstrated the expected dehydration and cyclization of the peptides (**Figure 5.2**). Subsequent removal of the leader peptides of modified $CylL_L$ or $CylL_S$ was achieved by CylA, the dedicated serine protease encoded by the cytolysin biosynthetic pathway (chapter 7). Electrospray ionization (ESI) MS analysis confirmed a mass shift of 144 or 126 Da for $CylL_L$, corresponding to a loss of 8 or 7 water molecules, with the 7-dehydrated peptide as the major product, and a mass shift of 72 Da for $CylL_S$, corresponding to a loss of 4 water molecules (**Figure 5.3**). These results are consistent with the reported mass of $CylL_L$" and $CylL_S$" as well as my previous observations using an *E. coli* coexpression system, where 7 dehydrations and 4 dehydrations were detected (chapter 2) (*8, 9*). Tandem MS (MS/MS) analysis indicated the desired ring systems were formed for both peptides (**Figure 5.4**).

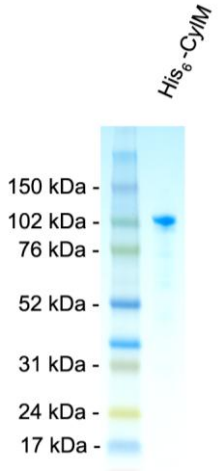

Figure 5.1 SDS-PAGE image of purified His_6-CylM. Theoretical molecular weight: 118 kDa.

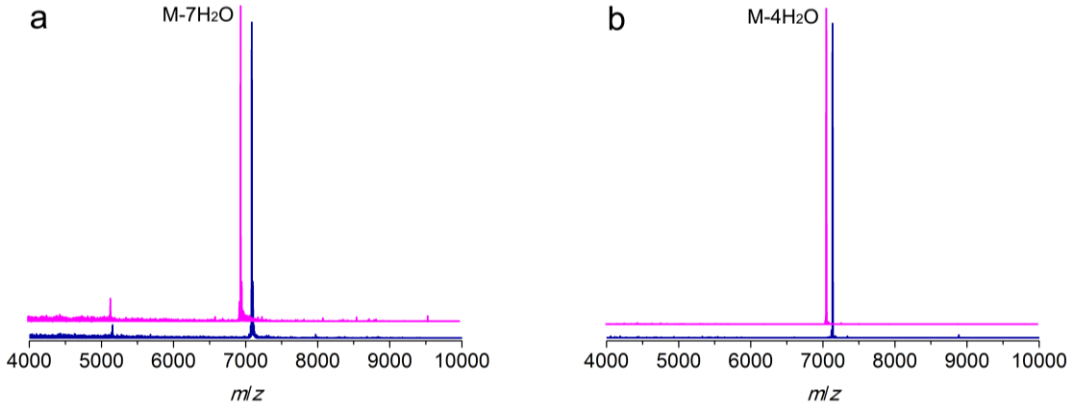

Figure 5.2 MALDI/TOF mass spectra for $CylL_L$ (a) and $CylL_S$ (b) peptides incubated with (magenta traces) or without (blue traces) CylM. Linear $CylL_L$, calculated M: 7,082, average mass; observed $M+H^+$: 7,084, average mass. CylM modified $CylL_L$, calculated $M–7H_2O$: 6,956, average mass; observed $M–7H_2O+H^+$: 6,958, average mass. $CylL_S$, calculated M: 7,132, average mass; observed $M+H^+$: 7,134, average mass. CylM modified $CylL_S$, calculated $M–4H_2O$: 7,060, average mass; observed $M–7H_2O+H^+$: 7,062, average mass.

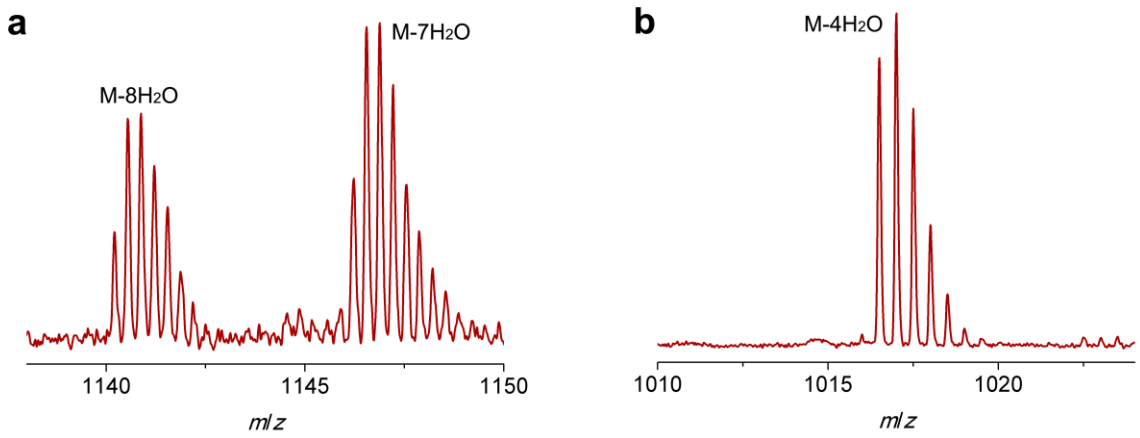

Figure 5.3 ESI/Q-TOF mass spectra for $CylL_L$ (a) and $CylL_S$ (b) core peptides modified by CylM *in vitro*. $[M+3H]^{3+}$ ion is shown for the $CylL_L$ core peptide and $[M+2H]^{2+}$ ion is shown for the $CylL_S$ core peptide. CylM modified $CylL_L$ core peptide, calculated $[M–7H_2O+3H]^{3+}$: 1146.2, monoisotopic mass; observed $[M–7H_2O+3H]^{3+}$: 1146.2, monoisotopic mass. CylM modified $CylL_S$ core peptide, calculated $[M–4H_2O+2H]^{2+}$: 1016.5, monoisotopic mass; observed $[M–4H_2O+2H]^{2+}$: 1016.5, monoisotopic mass.

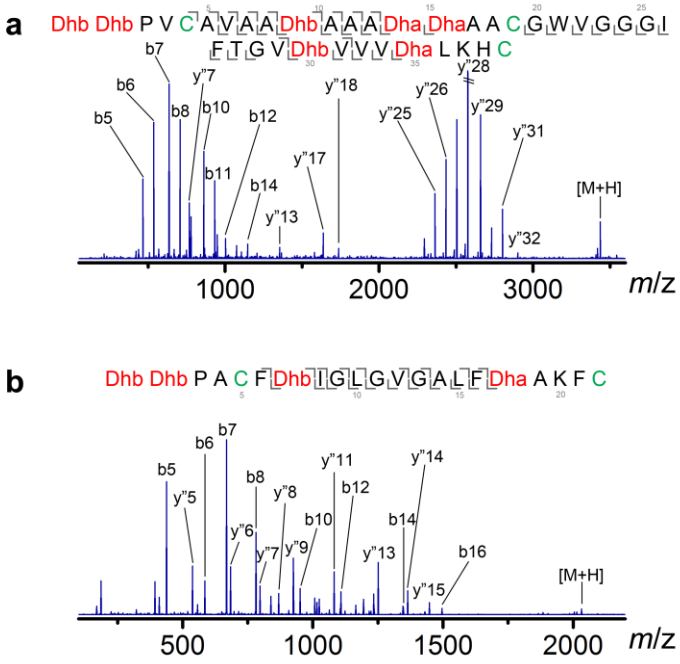

Figure 5.4 MS/MS analysis of $CylL_L$ (a) and $CylL_S$ (b) core peptides modified by CylM *in vitro*.

5.2.2 Structure of CylM

To further investigate the mechanism for both dehydration and cyclization afforded by a single enzyme, the 2.3 Å resolution structure of CylM, in complex with AMP, was determined by Dr. Shi-Hui Dong and Prof. Satish K. Nair. CylM consists of two distinct domains- an N-

terminal dehydration domain, composed of residues Asn4 through Pro624, and a C-terminal cyclization domain encompassed by Tyr641 through Glu992 (**Figure 5.5a**). Consistent with prior predictions, the cyclization domain is homologous to the stand-alone class I lanthionine cyclase NisC including the conserved zinc ligands (Cys875, Cys911, and His912) (**Figure 5.5b**) (*4*). The C-terminal cyclization domain of CylM contains a β-sheet region composed of three antiparallel strands (**Figures 5.5a** and **5.5c**). A similar antiparallel β-stranded element that interacts with the substrate leader peptide was identified in the class I lanthionine dehydratase NisB (*7*), and is a conserved feature in other enzymes involved in RiPP biosynthesis (**Figure 5.5c**) (*10*).

While class II dehydratases lack detectable sequence similarities with known proteins, the CylM structure reveals that this domain is architecturally similar to the catalytic core of phosphoinositide 3-kinase (PI3K) (**Figures 5.5a** and **5.5d**) (*11, 12*). The structural similarity with kinases is consistent with the proposed mechanism of dehydration via first phosphorylation of Ser and Thr residues, followed by elimination of the phosphate (*13*). Despite the structural homology, the topology of the CylM dehydration domain is quite different from those of canonical kinases (**Figures 5.5e**).

Like canonical kinases, the CylM dehydration domain is composed of an N-lobe and a C-lobe (**Figure 5.6a**). All protein and lipid kinases contain a requisite ~30-residue segment termed the activation loop that is important for both regulation and function. Kinase activity is typically controlled through activation-induced conformational changes, involving a disorder-to-order transition of the activation loop. Unlike most kinases, the activation loop in the CylM dehydration domain is well defined in the absence of bound peptide (**Figure 5.6a**). The stabilization of the activation loop is established through interactions with a LanM-specific KA domain (kinase-activation domain), which, is itself held in place by the "capping helices" (a two helical insert created by Ile180 through Tyr200) (**Figure 5.6a**). The observation that the activation loop stays in a catalytically competent conformation is consistent with the fact that the ATPase activity of LanM enzymes is not dependent on the supply of peptide substrate.

In the CylM-AMP cocrystal structure, the AMP is trapped in the cliff between the two lobes (**Figures 5.6a** and **5.6b**). The phosphate-binding loop (P-loop) is composed of residues Ser247 through Thr262, including Asp252 and His254 (**Figure 5.6b**). Mutational analysis suggests a second possible role for these residues in the elimination of phosphate from the phosphorylated

peptide intermediates (section 5.2.3). The adenine is situated in a hydrophobic pocket common across other structurally characterized kinases (*14, 15*). The proposed mechanism for protein kinases involves a conserved DxH motif, and LanM-conserved residues Asp347 and His349 in CylM are poised for catalysis (**Figure 5.6b**). His349 may serve as a general base and receive a proton from Ser/Thr in the substrate peptide. Asp347 likely orients the Ser/Thr oxygen of the substrate peptide for nucleophilic attack onto the γ-phosphate of ATP. Additionally, two conserved residues, Asn352 and Asp364, coordinate divalent metal ligands for stabilization of the incipient charge in the transition state. The CylM cocrystal structure further suggests a model for how the substrate peptide can bind to the two active sites (**Figures 5.5c** and **5.6c**).

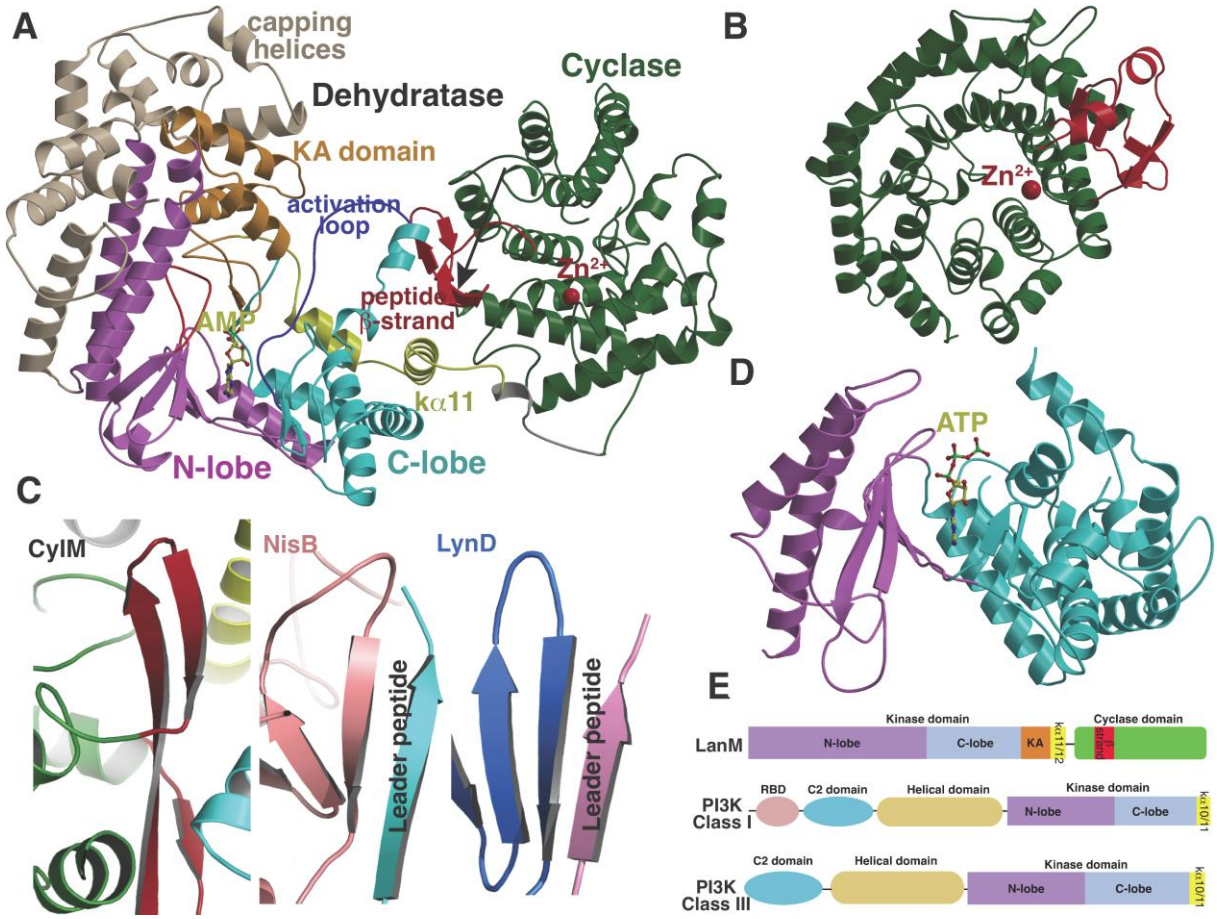

Figure 5.5 Crystal structure of CylM. (**a**) Overall structure of CylM. (**b**) Structure of the class I lanthipeptide cyclase NisC illustrating the structural homology with the C-terminus of CylM. (**c**) Comparison of the putative peptide-binding β-strands of CylM with the peptide binding regions of other RiPP biosynthetic enzymes including NisB (involved in nisin biosynthesis, PDB 4WD9) and LynD (involved in cyanobactin biosynthesis, PDB 4V1T). (**d**) Structure of the lipid kinase PI3K that shares homology with the dehydration domain of CylM. (**e**) Domain organization of LanMs in comparison with that of lipid kinases. RBD, Ras-binding domain. Figure made by Prof. Satish K. Nair.

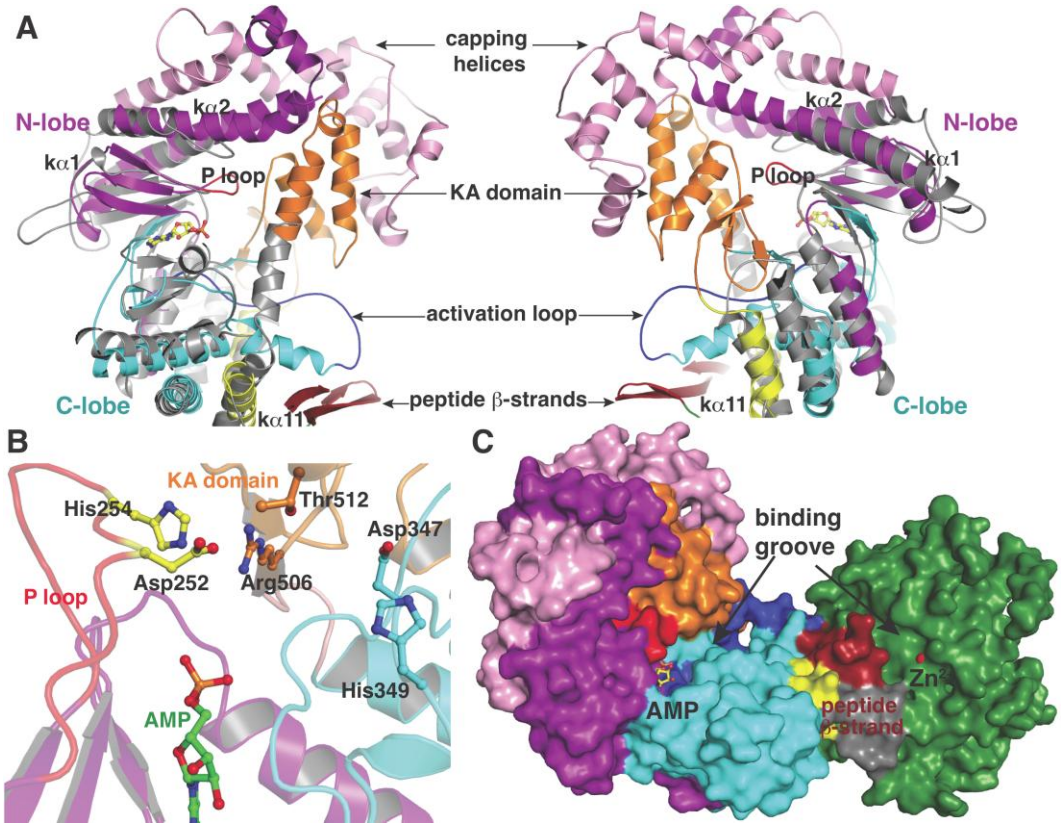

Figure 5.6 Active site geometry of CylM. (**a**) Two views, rotated by 180°, of the kinase active site of CylM showing the insertions specific to LanMs. Secondary structural elements are colored as in **Figure 5.5a** and the superimposed structure of PI3K is colored in gray. (**b**) Close up of the CylM dehydratase active site showing the bound nucleotide, and the residues that are important for phosphorylation and phosphate elimination. (**c**) Solvent occluded surface showing the roles of the capping helices and KA domain in orientating the activation loop. Figure made by Prof. Satish K. Nair.

5.2.3 Site-directed mutagenesis of active site residues

Given the unanticipated lipid kinase fold, I focused the mechanistic studies of CylM on the dehydratase domain (residues 1-624). The four residues in the ATP binding loop in CylM (Asp347, His349, Asn352, Asp364) are conserved in the LanM family (*16*). The importance of these residues was investigated by replacement with Ala in the *E. coli* coexpression system with $CylL_S$ as substrate. Dehydration activity was not detected for any of the mutants, with the exception of CylM-H349A that produced a small amount of dehydrated $CylL_S$ as determined by MALDI-TOF MS (**Figure 5.7**).

The activities of these CylM phosphorylation-deficient mutants were also tested *in vitro* using purified proteins (**Figure 5.8**). With linear $CylL_S$ serving as the substrate, wild type CylM

finished the modification by eliminating 4 water molecules within 30 min of incubation when characterized using MALDI-TOF MS. In comparison, the four phosphorylation-deficient mutants were unable to convert the starting material into modified peptide using the same set up (**Figure 5.9**). A small amount of dehydrated product was observed for CylM-H349A, which afforded partially dehydrated intermediates when analyzed using the *E. coli* coexpression system (**Figure 5.7**). I further increased the incubation time to 10 hours at room temperature to facilitate the detection of any minimal level activity. Indeed, almost full modification of $CylL_S$ was achieved by CylM-H349A (**Figure 5.10**). Partially modified products with one dehydration and one phosphorylation were observed for the reaction catalyzed by CylM-N352A, but $CylL_S$ remained unmodified in the presence of CylM-D347A and CylM-D364A even with elongated incubation time (**Figure 5.10**).

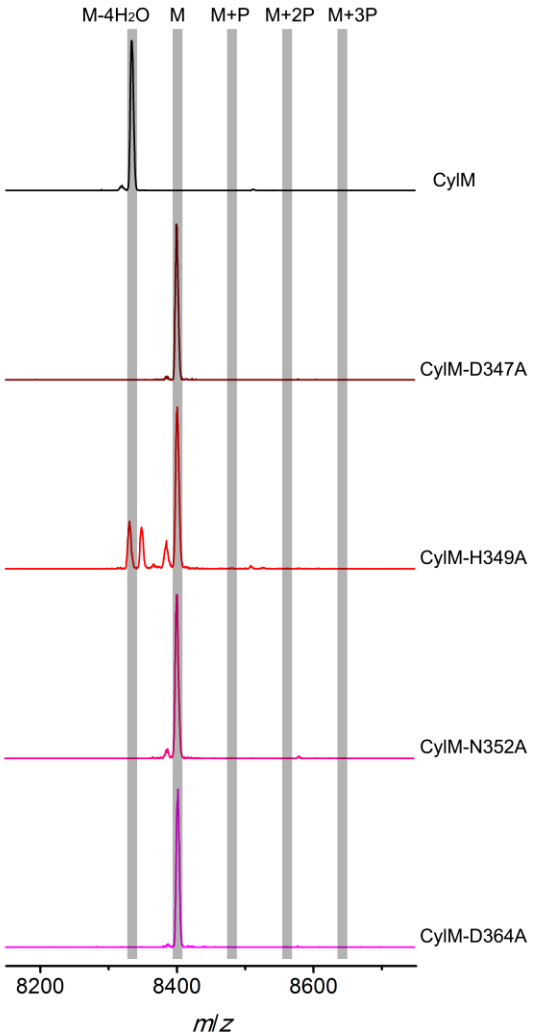

Figure 5.7 MALDI-TOF mass spectra of $CylL_S$ peptides coexpressed with CylM and CylM phosphorylation-deficient mutants in *E. coli*. M = unmodified $CylL_s$; P = phosphorylation.

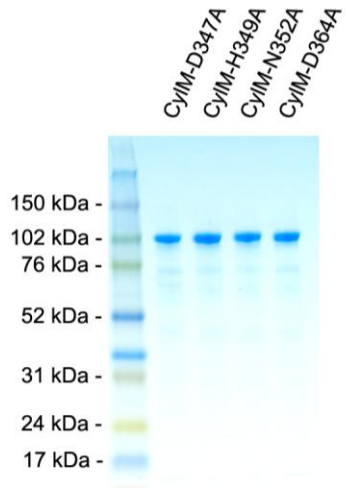

Figure 5.8 SDS-PAGE image of CylM phosphorylation-deficient mutants.

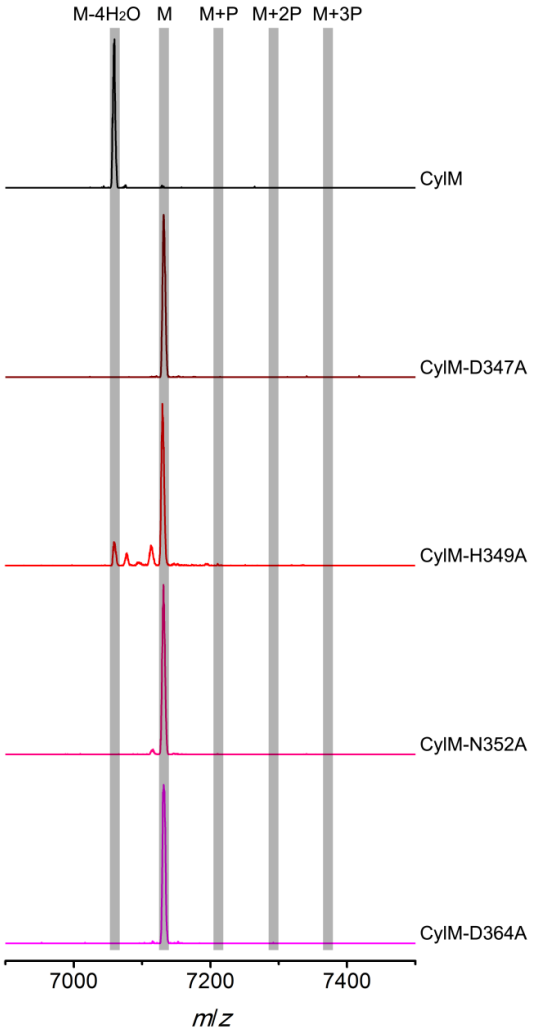

Figure 5.9 MALDI-TOF mass spectra of $CylL_S$ peptides incubated with CylM and CylM phosphorylation-deficient mutants *in vitro* for 30 min. M = unmodified $CylL_S$; P = phosphorylation.

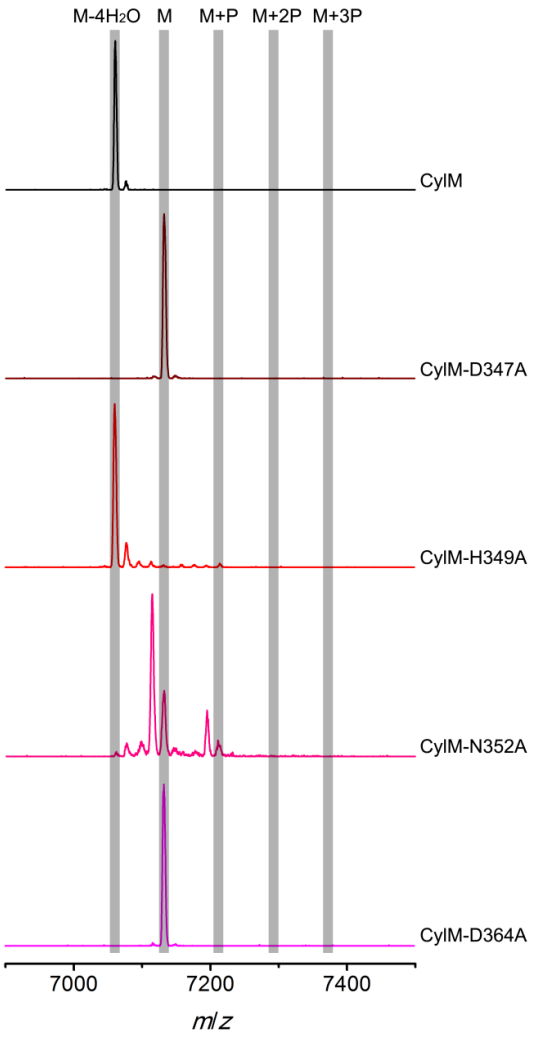

Figure 5.10 MALDI-TOF mass spectra of $CylL_S$ peptides incubated with CylM and CylM phosphorylation-deficient mutants *in vitro* for 10 hours. M = unmodified $CylL_s$; P = phosphorylation.

Previous mutagenesis studies identified four conserved residues in LanMs that are important for the elimination reaction (*13*, *17*). Inspection of the CylM structure shows that these four residues (Asp252, His254, Arg506, Thr512) are in close proximity despite being apart by > 250 amino acids in primary sequence (**Figure 5.6**) and are located adjacent to the phosphorylation site. The function of each residue was investigated using alanine substitution. Phosphorylated intermediates were detected for all four mutants, with partially dehydrated products observed for all except CylM-T512A (**Figure 5.11**). *In vitro* characterization of the four elimination-deficient mutants of CylM also provided similar phenotypes as what were observed using the coexpression system, except that CylM-H254A afforded fully modified $CylL_S$ after an elongated incubation period (**Figures 5.12-5.14**), indicating that mutating histidine 254 to alanine slows

down but does not abolish the phosphate-elimination activity of CylM. The T512A mutant did not eliminate the phosphate even with increased reaction time (**Figure 5.14**), suggesting that Thr512 is critical for the elimination activity of CylM.

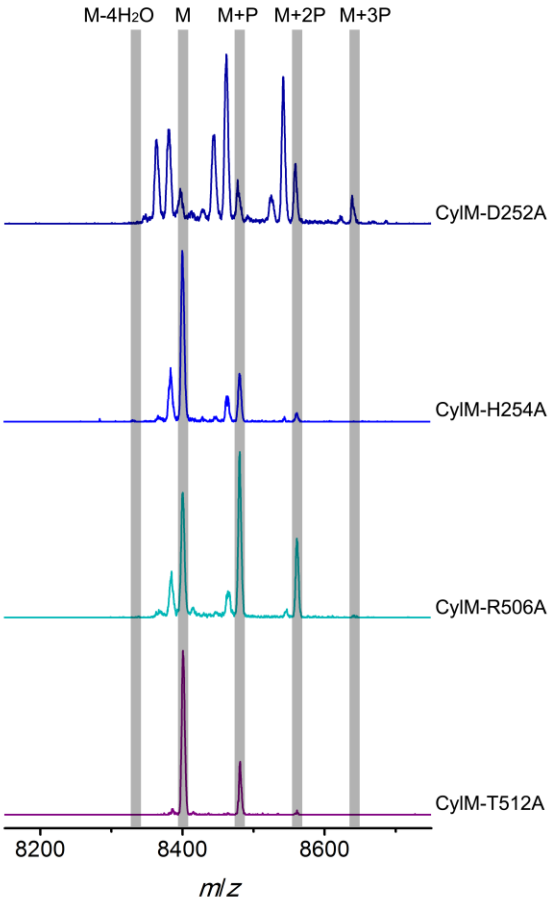

Figure 5.11 MALDI-TOF mass spectra of CylL$_S$ peptides coexpressed with CylM elimination-deficient mutants in *E. coli*. M = unmodified CylL$_s$; P = phosphorylation.

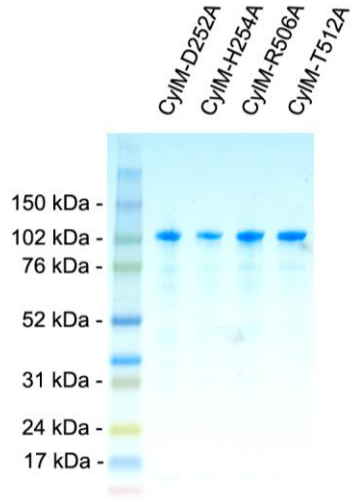

Figure 5.12 SDS-PAGE image of CylM elimination-deficient mutants.

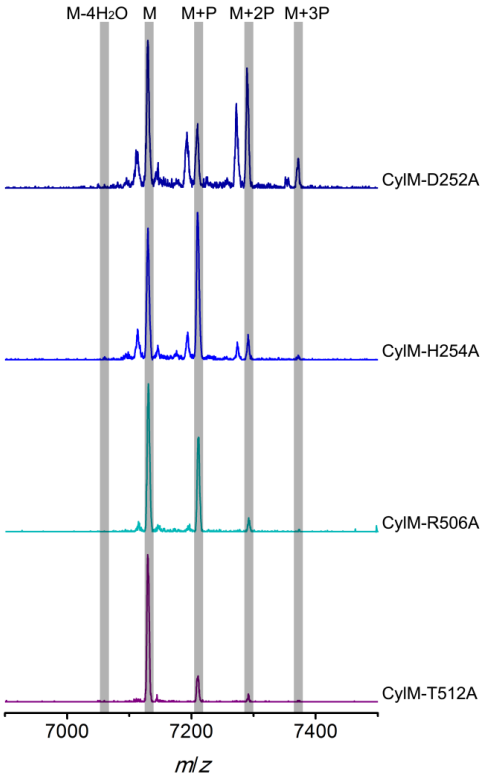

Figure 5.13 MALDI-TOF mass spectra of $CylL_S$ peptides incubated with CylM elimination-deficient mutants *in vitro* for 30 min. M = unmodified $CylL_s$; P = phosphorylation.

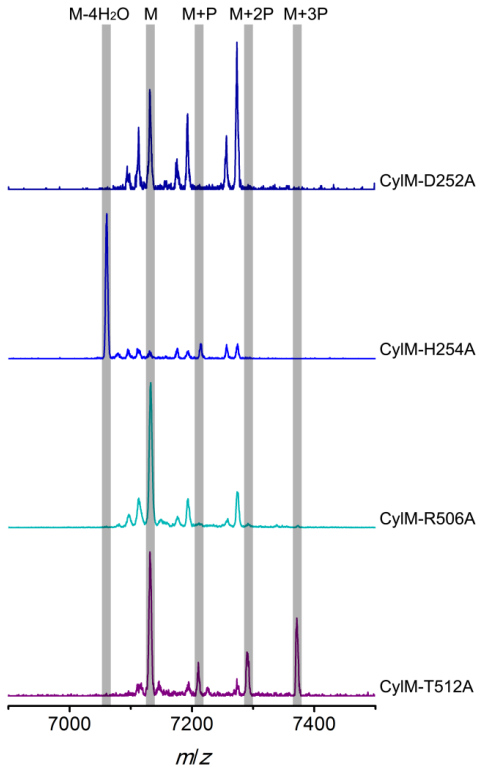

Figure 5.14 MALDI-TOF mass spectra of $CylL_S$ peptides incubated with CylM elimination-deficient mutants *in vitro* for 10 hours. M = unmodified $CylL_s$; P = phosphorylation.

Both Asp252 and His254 are residues within the kinase domain P loop, which is thought to facilitate ATP activation by interacting with the γ phosphate, and it is of interest to note that these residues may play a similar role of phosphate stabilization during the elimination reaction. Lastly, both Arg506 and Thr512 are located within the KA domain, suggesting that, in addition to stabilizing LanMs in an active conformation, this domain may also provide residues to assist in the elimination of phosphate from the phosphorylated peptide intermediate.

5.2.4 ADP is required for CylM-catalyzed phosphate elimination

Unlike the class III lanthipeptide synthetases for which phosphorylated intermediates accumulate (*18*), such intermediates are usually not detected during catalysis by LanM proteins. Class III synthetases are made up of separate protein kinase, phosphoSer/phosphoThr elimination, and cyclase domains that are readily recognized by sequence homology (*6*). The two distinct domains for dehydration require the phosphorylated peptides to translocate from the kinase to the lyase active site. The adjacency of the phosphorylation and elimination active sites in CylM provides an explanation for the absence of phosphorylated peptides in LanM proteins as elimination may occur faster than peptide dissociation. To further investigate the elimination step, a mixture of phosphorylated $CylL_S$ peptides carrying different numbers of phosphate esters was obtained by coexpression of $CylL_S$ with elimination-deficient CylM-R506A in *E. coli*. The phosphorylated $CylL_S$ peptides were incubated with wild type CylM in the presence or absence of adenosine nucleotides. Without addition of nucleotides, CylM did not eliminate the phosphates. However, when ADP was supplied, phosphorylated $CylL_S$ was consumed, resulting in a mixture of $CylL_S$ peptides that differed in the number of dehydrations (**Figures 5.15** and **5.16**). Collectively, these results are consistent with an ordered kinetic mechanism in which binding of the phosphorylated peptide requires ADP to be bound first or in which the presence of ADP within the active site increases the affinity of phosphorylated intermediates.

5.2.5 Conclusion

The unexpected structural homology of bacterial LanM proteins with eukaryotic lipid kinases and related protein kinases may have implications for the function of the three LanC-like proteins that are found in mammals (*19-21*). Recent studies indicated interactions between the human LanCL2 and the kinases Akt and the mechanistic target of rapamycin complex 2

(mTORC2) (*22*). The structure of CylM shows that the LanC domain interacts with the activation loop and the kα11 helix of the kinase domain, providing a potential preview of the intermolecular interaction of LanCL proteins with mammalian kinases.

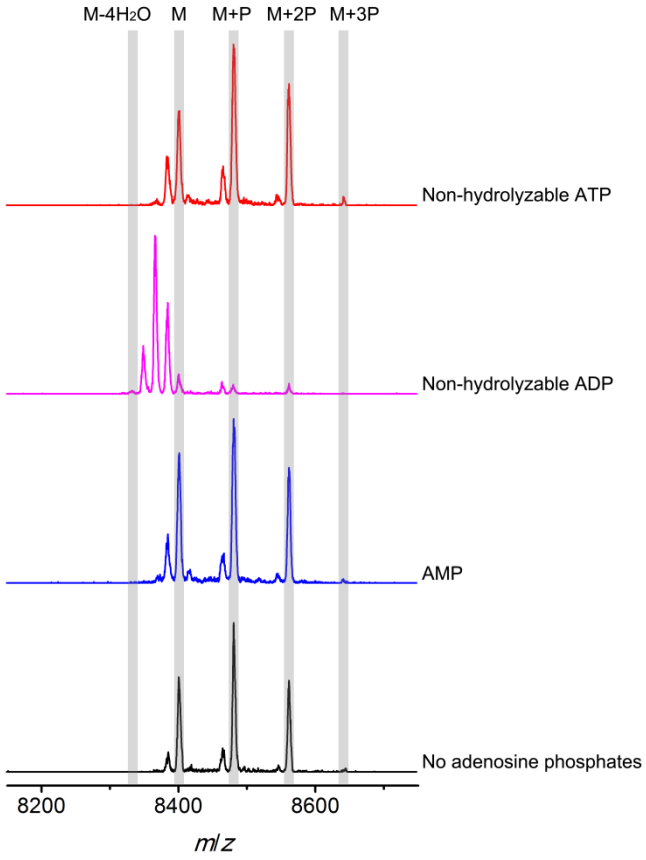

Figure 5.15 MALDI-TOF mass spectra of phosphorylated $CylL_S$ intermediates incubated with CylM in the absence of nucleotides (black trace), and in the presence of AMP (adenosine 5'-monophosphate disodium salt) (blue trace), non-hydrolyzable ADP (adenosine 5'-(β-thio)diphosphate trilithium salt) (magenta trace), or non-hydrolyzable ATP (adenosine 5'-(β,γ-imido)triphosphate lithium salt hydrate) (red trace). M = unmodified $CylL_s$; P = phosphorylation. The data are shown for non-hydrolyzable analogs of ADP and ATP to distinguish whether the observed activity is due to the presence of these nucleotides, or to the activated phosphor-anhydride groups of ADP/ATP. See also **Figure 5.16**.

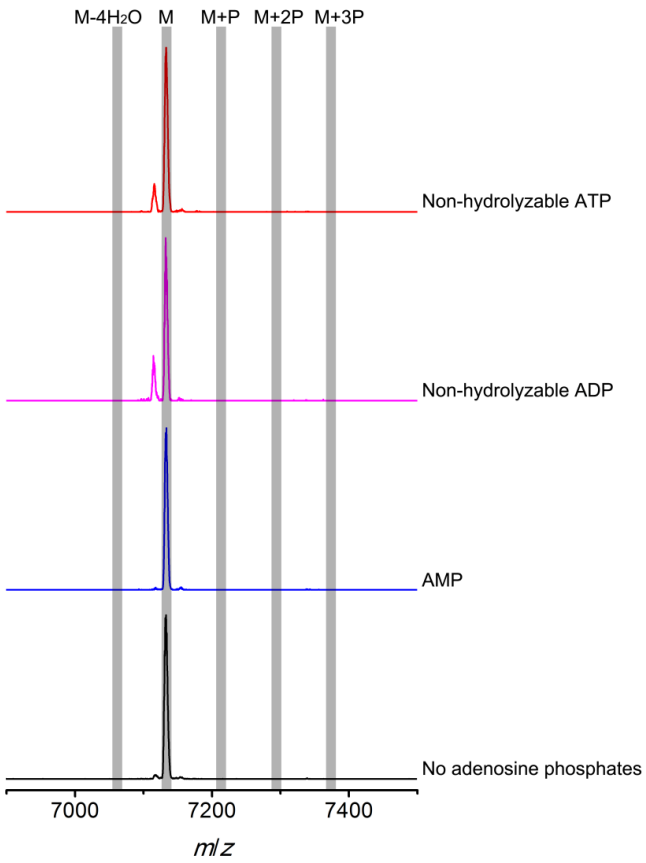

Figure 5.16 MALDI-TOF mass spectra of $CylL_S$ peptides incubated with CylM in the absence of nucleotides (black trace), and in the presence of AMP (adenosine 5'-monophosphate disodium salt) (blue trace), non-hydrolyzable ADP (adenosine 5'-(β-thio)diphosphate trilithium salt) (magenta trace), or non-hydrolyzable ATP(adenosine 5'-(β,γ-imido)triphosphate lithium salt hydrate) (red trace). M = unmodified $CylL_S$; P = phosphorylation.

5.3 Methods

General methods

Similar general methods and materials as what were described in chapter 2 were employed in this chapter unless specified otherwise. LC-ESI-Q/TOF MS analyses were conducted using a Synapt G2 MS system equipped with Acquity UPLC (Waters). Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) was carried out on a Bruker Daltonics UltrafleXtreme MALDI-TOF/TOF mass spectrometer. C18 zip-tip pipet tips were obtained from Millipore to desalt samples for MS analysis. Chemicals were ordered from Sigma Aldrich. Expression vectors (pET15b and pRSFDuet-1) were obtained from Novagen.

Cloning of *cylM*, *cylL$_L$*, and *cylL$_S$* genes into expression vectors

The c*ylM* gene was cloned into the multiple cloning site 1 of a pRSFDuet-1 vector using *Eco*RI and *Not*I restriction sites to generate pRSFDuet-1/CylM plasmid. Primer sequences used are listed in **Table 5.1**. C*ylL$_L$* and *cylL$_S$* genes were cloned into a pET15b vector using *Nde*I and *BamH*I restriction sites, resulting in pET15b/CylL$_L$ or pET15b/CylL$_S$ plasmids, respectively.

Table 5.1 Primer sequences used for cloning of *cylM* and its mutants.

Primer Name	Primer Sequence (5'-3')
CylM_EcoRI_Duet_FP	AAAAA GAATTCG GAAGATA ATCTGATTAA T
CylM_NotI_Duet_RP	AAAAA GCGGCCGC TTACAGT TCAAACAGCA G
CylM_D252A_QC_FP	AGGGT GCA AGCCAT AGCCGTGGTAAAACCGTT AGC
CylM_D252A_QC_RP	ATGGCT TGC ACCCT GGC TTTCGCTAAT GCTATTCAGT
CylM_H254A_QC_FP	GATAGC GCT AGCCGT GGT AAAACCGTT AGCACCCTG
CylM_H254A_QC_RP	ACGGCT AGC GCTA TC ACCCTGGC TTTCGCTAAT G
CylM_D347A_QC_FP	GTTACC GCT CTGCAT TATGAAAACATCATTGCCCATGGC
CylM_D347A_QC_RP	AT GCAG AGC GGT AAC ATTAAAC AGAAAGGCAA TGCCAATCAG
CylM_H349A_QC_FP	CCGATCTG GCT TATGAAAA CATCATTGCCCATGGCGAATA
CylM_H349A_QC_RP	TTTTCATA AGC CAGATCGG T AACATTAAAC AGAAAGGCAA TGCCAAT
CylM_N352A_QC_FP	CATTATGAA GCC ATCATTGC CCATGGCGAATATCCG GTGATT
CylM_N352A_QC_RP	GCAATGAT GGC TTCATAATG CAGATCGGT AACATTAAAC AGAAAGGC
CylM_D364A_QC_FP	GTGATTATT GCT AATGAAACC TTTTTTCAGCAGAATATTCCGATTGAATTT
CylM_D364A_QC_RP	GGTTTC ATT AGC AATA ATCAC CGGAT ATTCGCCATG GGC
CylM_R506A_QC_FP	TGATTGTG GCC AATGTTAT TCGTCCGACCCAGCGTTA
CylM_R506A_QC_RP	A TAACATT GGC CACAATCA GA TTCTGCAGAT TATTATTAAT ATAGGCCAGA
CylM_T512A_QC_FP	GTCCG GCC CAG C GTTATGCAGATATGCTGGAA TTTAGC
CylM_T512A_QC_RP	CTG GGCCGGAC GAA TAACATTGCG CACAATCAGA

Construction of pRSFDuet-1 derivatives for expression of CylM mutants and for coexpression of CylL$_S$ with CylM mutants

The plasmids pRSFDuet-1/CylM-D347A, pRSFDuet-1/CylM-H349A, pRSFDuet-1/CylM-N352A, pRSFDuet-1/CylM-D364A, pRSFDuet-1/CylM-D252A, pRSFDuet-1/CylM-H254A, pRSFDuet-1/CylM-R506A, pRSFDuet-1/CylM-T512A, pRSFDuet-1/CylL$_S$/CylM-D347A-2, pRSFDuet-1/CylL$_S$/CylM-H349A-2, pRSFDuet-1/CylL$_S$/CylM-N352A-2, pRSFDuet-

1/CylL_S/CylM-D364A-2, pRSFDuet-1/CylL_S/CylM-D252A-2, pRSFDuet-1/CylL_S/CylM-H254A-2, pRSFDuet-1/CylL_S/CylM-R506A-2 and pRSFDuet-1/CylL_S/CylM-T512A-2 were generated using QuikChange methodology using pRSFDuet-1/CylM and pRSFDuet-1/CylL_S/CylM-2 as templates, respectively (*8*). Primer sequences are listed in **Table 5.1**.

Expression and purification of His_6-CylL_L and His_6-CylL_S peptides

E. coli BL21 (DE3) cells were transformed with pET15b/CylL_L or pET15b/CylL_S and plated on a LB plate containing 100 mg/L ampicillin. A single colony was picked and grown in 20 mL of LB in the presence of ampicillin at 37 ^oC for 12 h. The cell suspension was directly used to inoculate 2 L of fresh LB media. Cells were cultured at 37 ^oC until the OD at 600 nm reached 0.5, and IPTG was added to a final concentration of 0.2 mM. Cells were cultured at 37 ^oC for another 3-5 h before harvesting. The cell pellet was resuspended at room temperature in LanA start buffer (20 mM NaH_2PO_4, 500 mM NaCl, 0.5 mM imidazole, 20% glycerol, pH 7.5 at 25 ^oC) and lysed by sonication. The resulting sample was then centrifuged at 23,700 ×g for 30 min and supernatant was discarded. The remaining pellet was resuspended in LanA buffer 1 (6 M guanidine hydrochloride, 20 mM NaH_2PO_4, 500 mM NaCl, 0.5 mM imidazole, pH 7.5 at 25 ^oC) and sonicated. Centrifugation was performed afterwards to pellet the debris and the soluble portion was passed through 0.45-μm syringe filters. His-tagged peptides were purified by immobilized metal ion affinity chromatography (IMAC) eluting with LanA elute buffer (4 M guanidine hydrochloride, 20 mM NaH_2PO_4, 500 mM NaCl, 1 M imidazole , pH 7.5 at 25 ^oC). The eluted fractions were desalted by preparative HPLC using a Waters Delta-pak C4 column (15 μm 300 Å 25 x 100 mm). The resulting peptides were lyophilized to dryness and kept at −20 ^oC until future use.

Expression and purification of His_6-CylM and CylM mutants

E. coli BL21 (DE3) cells were transformed with pRSFDuet-1/CylM, pRSFDuet-1/CylM-D347A, pRSFDuet-1/CylM-H349A, pRSFDuet-1/CylM-N352A, pRSFDuet-1/CylM-D364A, pRSFDuet-1/CylM-D252A, pRSFDuet-1/CylM-H254A, pRSFDuet-1/CylM-R506A or pRSFDuet-1/CylM-T512A, and plated on a LB plate containing 50 mg/L kanamycin. A single colony was picked and grown in 20 mL of LB in the presence of kanamycin at 37 ^oC for 12 h. The cell suspension was directly used to inoculate 2 L of LB and cells were cultured at 37 ^oC

until the OD at 600 nm reached 0.5. The culture was cooled down on ice followed by the addition of IPTG to a final concentration of 0.1 mM. Cells were cultured at 18 ^{o}C for additional 18 h before harvesting. The harvested cells were resuspended on ice in LanM start buffer (20 mM HEPES, 1 M NaCl, pH 7.5 at 25 ^{o}C) and lysed using a homogenizer. Insoluble debris was removed by centrifugation at 23,700 ×g for 45 min at 4 ^{o}C and the supernatant was passed through 0.45-μm syringe filters. His-tagged proteins were purified by IMAC, eluting with a linear concentration gradient of imidazole from 30 mM to 200 mM. The eluted fractions were analyzed using SDS-PAGE. Fractions containing the desired protein were combined and concentrated using a centrifugal filtering device, and the buffer was exchanged to LanM start buffer using a gel-filtration column. Protein concentration was quantified by its absorbance at 280 nm. The extinction coefficient for His_6-CylM was calculated as 140,110 M^{-1} cm^{-1}. Aliquoted protein solutions were flash-frozen and kept at −80 ^{o}C until further usage.

Expression and purification of His_6-$CylL_S$ peptides coexpressed with CylM mutants

E. coli BL21 (DE3) cells were transformed with pRSFDuet-1/$CylL_S$/CylM-D347A-2, pRSFDuet-1/$CylL_S$/CylM-H349A-2, pRSFDuet-1/$CylL_S$/CylM-N352A-2, pRSFDuet-1/$CylL_S$/CylM-D364A-2, pRSFDuet-1/$CylL_S$/CylM-D252A-2, pRSFDuet-1/$CylL_S$/CylM-H254A-2, pRSFDuet-1/$CylL_S$/CylM-R506A-2 or pRSFDuet-1/$CylL_S$/CylM-T512A-2, and plated on a LB plate containing 50 mg/L kanamycin. A single colony was picked and grown in 10 mL of LB in the presence of kanamycin at 37 ^{o}C for 12 h. The cell suspension was directly used to inoculate 1 L of LB and cells were cultured at 37 ^{o}C until the OD at 600 nm reached 0.5. The culture was cooled down on ice followed by the addition of IPTG to a final concentration of 0.1 mM. Cells were cultured at 18 ^{o}C for 18 h before harvesting. To obtain both fully modified and linear $CylL_S$ as well as possible intermediates (partially modified $CylL_S$) and reduce the bias introduced by peptide solubility, harvested cells were resuspended and lysed directly in LanA buffer 1 (6 M guanidine hydrochloride, 20 mM NaH_2PO_4, 500 mM NaCl, 0.5 mM imidazole, pH 7.5 at 25 ^{o}C) by sonication. Debris was removed by centrifugation and the soluble portion was passed through 0.45-μm syringe filters. His-tagged $CylL_S$ was purified by immobilized metal ion affinity chromatography (IMAC), eluting with LanA elute buffer (4 M guanidine hydrochloride, 20 mM NaH_2PO_4, 500 mM NaCl, 1 M imidazole, pH 7.5 at 25 ^{o}C). The eluted fractions were desalted with Strata-X polymeric reverse phase SPE columns and lyophilized to dryness.

Reconstitution of CylM activity *in vitro*

Presumably due to high hydrophobicity, the solubility of linear $CylL_L$ or $CylL_S$ peptides is extremely poor. For enzyme assays, 2 mg/mL peptide suspension was made in deionized water as stock solution for both peptides. The stock solution was vortexed to a homogenized suspension each time before any peptide was taken. However, given the presence of precipitation, the concentration of $CylL_L$ or $CylL_S$ peptides could not be tightly controlled.

To reconstitute the activity of CylM *in vitro*, 20 μM of linear peptides were supplied in a reaction vessel with 4 mM $MgCl_2$, 2 mM ATP, 2 mM DTT, 1 x 10^{-5} U thrombin (to remove the His-tag in situ) and 50 mM HEPES (pH 7.5), followed by the addition of CylM to a final concentration of 0.5 μM. Reactions were incubated at room temperature for 4 h. Control reactions were set up with all other components in the absence of CylM. Each sample was zip-tipped and analyzed by MALDI-TOF MS. Aliquoted samples were treated by CylA (serine protease encoded in the biosynthetic pathway of cytolysin) to remove the leader peptides, and the resulting core peptides were analyzed by LC-MS or LC-MS/MS.

***In vitro* modification of $CylL_S$ by CylM mutants**

$CylL_S$ peptide (20 μM) was supplied to a reaction vessel in the presence of 4 mM $MgCl_2$, 2 mM ATP, 2 mM DTT, 1 x 10^{-5} U thrombin (to remove the His-tag in situ) and 50 mM HEPES (pH 7.5). CylM and CylM mutant proteins were then added to a final concentration of 0.5 μM. Reactions were incubated at room temperature and aliquots were quenched by adding formic acid to a final concentration of 0.5% at desired time points. Each sample was then zip-tipped and analyzed by MALDI-TOF MS.

Elimination activity of CylM in the presence of adenosine derivatives

Phosphorylated $CylL_S$ peptides carrying different numbers of phosphate esters were obtained by coexpression of $CylL_S$ with CylM elimination-deficient mutant CylM-R506A. The peptide mixture was dissolved in deionized water to make a 350 μM stock solution. Non-hydrolyzable ATP (adenosine 5'-(β,γ-imido)triphosphate lithium salt hydrate), non-hydrolyzable ADP (adenosine 5'-(β-thio)diphosphate trilithium salt) and AMP (adenosine 5'-monophosphate disodium salt) were reconstituted in deionized water and a stock solution of 20 mM was obtained

for each. For elimination reactions, CylM was supplied with a final concentration of 0.5 μM in the presence of 1 mM $MgCl_2$, 2 mM DTT and 50 mM HEPES (pH 7.5) in a reaction vessel, followed by the addition of adenosine derivatives (final concentration 500 μM) or deionized water (negative control). Phosphorylated $CylL_S$ peptides were then added into the mixture to a final concentration of 35 μM and incubated at room temperature for 2 h. Parallel reactions were set up using linear $CylL_S$ with a peptide concentration of 20 μM in the presence of 1 x 10^{-5} U thrombin. Samples were zip-tipped and analyzed by MALDI-TOF MS.

Crystallization and structural determination of CylM

This part was performed by Dr. Shi-Hui Dong and Prof. Satish K. Nair.

5.4 References

1. Gilmore, M. S., Segarra, R. A., Booth, M. C., Bogie, C. P., Hall, L. R., and Clewell, D. B. (1994) Genetic structure of the Enterococcus faecalis plasmid pAD1-encoded cytolytic toxin system and its relationship to lantibiotic determinants, *J. Bacteriol. 176*, 7335-7344.
2. Cox, C. R., Coburn, P. S., and Gilmore, M. S. (2005) Enterococcal cytolysin: a novel two component peptide system that serves as a bacterial defense against eukaryotic and prokaryotic cells, *Curr. Protein Pept. Sci. 6*, 77-84.
3. Zhang, Q., Yu, Y., Velásquez, J. E., and van der Donk, W. A. (2012) Evolution of lanthipeptide synthetases, *Proc. Natl. Acad. Sci. U. S. A. 109*, 18361-18366.
4. Li, B., Yu, J. P., Brunzelle, J. S., Moll, G. N., van der Donk, W. A., and Nair, S. K. (2006) Structure and mechanism of the lantibiotic cyclase involved in nisin biosynthesis, *Science 311*, 1464-1467.
5. Li, H., Xu, H., Zhou, Y., Zhang, J., Long, C., Li, S., Chen, S., Zhou, J. M., and Shao, F. (2007) The phosphothreonine lyase activity of a bacterial type III effector family, *Science 315*, 1000-1003.
6. Goto, Y., Li, B., Claesen, J., Shi, Y., Bibb, M. J., and van der Donk, W. A. (2010) Discovery of unique lanthionine synthetases reveals new mechanistic and evolutionary insights, *PLoS Biol. 8*, e1000339.
7. Ortega, M. A., Hao, Y., Zhang, Q., Walker, M. C., van der Donk, W. A., and Nair, S. K. (2015) Structure and mechanism of the tRNA-dependent lantibiotic dehydratase NisB, *Nature 517*, 509-512.
8. Tang, W., and van der Donk, W. A. (2013) The sequence of the enterococcal cytolysin imparts unusual lanthionine stereochemistry, *Nat. Chem. Biol. 9*, 157-159.
9. Booth, M. C., Bogie, C. P., Sahl, H.-G., Siezen, R. J., Hatter, K. L., and Gilmore, M. S. (1996) Structural analysis and proteolytic activation of Enterococcus faecalis cytolysin, a novel lantibiotic, *Mol. Microbiol. 21*, 1175-1184.
10. Koehnke, J., Bent, A. F., Zollman, D., Smith, K., Houssen, W. E., Zhu, X., Mann, G., Lebl, T., Scharff, R., Shirran, S., Botting, C. H., Jaspars, M., Schwarz-Linek, U., and Naismith, J. H. (2013) The cyanobactin heterocyclase enzyme: a processive adenylase that operates with a defined order of reaction, *Angew. Chem. Int. Ed. Engl. 52*, 13991-13996.
11. Williams, R., Berndt, A., Miller, S., Hon, W. C., and Zhang, X. (2009) Form and flexibility in phosphoinositide 3-kinases, *Biochem. Soc. Trans. 37*, 615-626.
12. Walker, E. H., Perisic, O., Ried, C., Stephens, L., and Williams, R. L. (1999) Structural insights into phosphoinositide 3-kinase catalysis and signalling, *Nature 402*, 313-320.
13. You, Y. O., and van der Donk, W. A. (2007) Mechanistic investigations of the dehydration reaction of lacticin 481 synthetase using site-directed mutagenesis, *Biochemistry 46*, 5991-6000.

14. Bao, Z. Q., Jacobsen, D. M., and Young, M. A. (2011) Briefly bound to activate: transient binding of a second catalytic magnesium activates the structure and dynamics of CDK2 kinase for catalysis, *Structure 19*, 675-690.
15. Yang, H., Rudge, D. G., Koos, J. D., Vaidialingam, B., Yang, H. J., and Pavletich, N. P. (2013) mTOR kinase structure, mechanism and regulation, *Nature 497*, 217-223.
16. Miller, S., Tavshanjian, B., Oleksy, A., Perisic, O., Houseman, B. T., Shokat, K. M., and Williams, R. L. (2010) Shaping development of autophagy inhibitors with the structure of the lipid kinase Vps34, *Science 327*, 1638-1642.
17. Ma, H., Gao, Y., Zhao, F., Wang, J., Teng, K., Zhang, J., and Zhong, J. (2014) Dissecting the catalytic and substrate binding activity of a class II lanthipeptide synthetase BovM, *Biochem. Biophys. Res. Commun. 450*, 1126-1132.
18. Jungmann, N. A., Krawczyk, B., Tietzmann, M., Ensle, P., and Sussmuth, R. D. (2014) Dissecting reactions of nonlinear precursor peptide processing of the class III lanthipeptide curvopeptin, *J. Am. Chem. Soc. 136*, 15222-15228.
19. Zhong, W. X., Wang, Y. B., Peng, L., Ge, X. Z., Zhang, J., Liu, S. S., Zhang, X. N., Xu, Z. H., Chen, Z., and Luo, J. H. (2012) Lanthionine synthetase C-like protein 1 interacts with and inhibits cystathionine beta-synthase: a target for neuronal antioxidant defense, *J. Biol. Chem. 287*, 34189-34201.
20. Huang, C., Chen, M., Pang, D., Bi, D., Zou, Y., Xia, X., Yang, W., Luo, L., Deng, R., Tan, H., Zhou, L., Yu, S., Guo, L., Du, X., Cui, Y., Hu, J., Mao, Q., Worley, P. F., and Xiao, B. (2014) Developmental and Activity-Dependent Expression of LanCL1 Confers Antioxidant Activity Required for Neuronal Survival, *Dev. Cell 30*, 479-487.
21. Chung, C. H. Y., Kurien, B. T., Mehta, P., Mhatre, M., Mou, S., Pye, Q. N., Stewart, C., West, M., Williamson, K. S., Post, J., Liu, L., Wang, R., and Hensley, K. (2007) Identification of lanthionine synthase C-like protein-1 as a prominent glutathione binding protein expressed in the mammalian central nervous system, *Biochemistry 46*, 3262-3269.
22. Zeng, M., van der Donk, W. A., and Chen, J. (2014) Lanthionine synthetase C-like protein 2 (LanCL2) is a novel regulator of Akt, *Mol. Biol. Cell 25*, 3954-3961.

Chapter 6. Class II Lanthipeptides Harbor a Pool of Sequence-Specific LanP Proteases

6.1 Introduction

Proteases are ubiquitous in most organisms (*1*). They have been the focus of intense research not only because of their pivotal physiological functions but also because of their potential applications. The use of proteases has led to advances in analytical chemistry, proteomics, medicine, and the food, detergent and leather industries (*2-4*). Although much effort has been spent on engineering of proteases with desired sequence specificity using both rational design and high throughput screening (*2, 5, 6*), few such efforts have reached the stage of commercial applications (*3*). Major challenges are the sacrifice of efficiency and stability when engineering new substrate specificity, or the loss of sequence specificity when focusing on improving protein robustness (*7-9*). As a result, nature is still the major source of proteases with novel recognition sequences. The biosynthetic machinery responsible for the production of RiPPs (*10*) is a promising area for discovering new sequence-specific proteases, as dedicated proteolytic enzymes are employed to remove leader peptides with highly diverse P and P' positions. Lanthipeptides are a large family of RiPPs with a subfamily that exhibit antimicrobial activities termed lantibiotics (for an introduction of lanthipeptides and lantibiotics, see section 1.2 in chapter 1).

Lanthipeptides are currently classified in four classes according to the synthetases that install the thioether cross-linked amino acids (**Figure 1.3a**) (*11*). Compared to the well-characterized lanthionine synthetases, the proteases responsible for leader peptide removal are much less studied (*12-19*). Two types of proteases have been reported for the maturation of class I and II lanthipeptides (**Figure 1.4a**) - the subtilisin-like serine protease LanP employed by both classes and the papain-like cysteine protease domain of the LanT transporter protein involved exclusively in the biosynthesis of class II lanthipeptides (*12-18*). The cysteine protease domain is located at the N-terminus of LanT proteins and typically cleaves its substrate after a double Gly-type motif (GG/GA/GS) (*15, 16*). In contrast to LanT protease domains, the sequence specificity of subtilisin-like LanP proteins remains mostly elusive. Thus far, only three class I LanP proteases have been heterologously expressed and characterized *in vitro* (*14, 20-23*), whereas no such studies have been performed for class II LanPs, which fall into a different phylogenetic clade (*23*).

The biosynthesis of only a few known class II lanthipeptides involves LanP proteases (*11*). These enzymes remove a short N-terminal oligopeptide after a LanT protein detaches the majority of the leader peptide at a double Gly-type cleavage site. For example, LicP is an extracellularly located serine protease expressed by some strains of *Bacillus licheniformis* and is required for the production of the two-component lantibiotic lichenicidin (**Figure 6.1**) (*18*). After installation of the thioether rings in the precursor peptides LicA1 and LicA2, LicT removes the leader peptide of modified LicA1 to generate Licα as well as the majority of the leader peptide from modified LicA2 to generate NDVNPE-Licβ (hereafter LicA2') (**Figure 6.2**) (*18*). The maturation of Licβ requires one more cleavage step outside the cell, where LicP trims off the six remaining amino acids at the N-terminus of LicA2' (**Figure 6.2**).

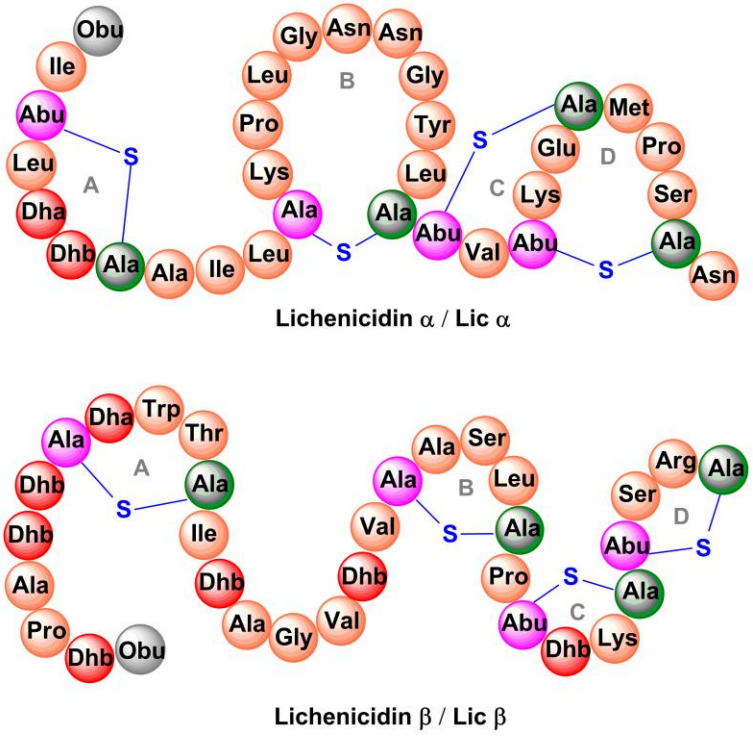

Figure 6.1 Structure of the lantibiotic lichenicidin. Obu, 2-oxobutyryl group resulting from hydrolysis of an N-terminal Dhb; Abu, α-aminobutyric acid.

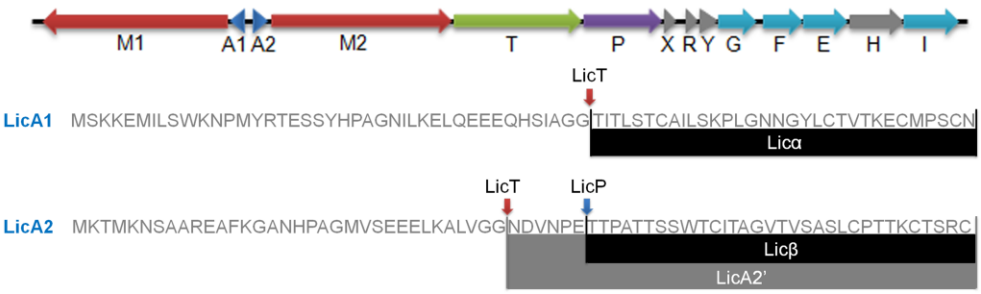

Figure 6.2 The biosynthetic gene cluster of lichenicidin and the cleavage events employed during lichenicidin maturation.

In this study, I describe the *in vitro* characterization of LicP, demonstrate that the protein self-cleaves, and show that it recognizes a specific cleavage sequence but otherwise is very tolerant of the peptide sequence. In addition, by collaborating with Dr. Shi-Hui Dong and Prof. Satish Nair, a 2.35 Å resolution structure of LicP was obtained, which reveals an unanticipated intramolecular activation strategy that stabilizes the mature form of the enzyme in a calcium-independent manner. LicP has potential utility for sequence-specific proteolysis because it does not leave a scar (*3, 24*). Encouraged by these observations, I identified nine new class II LanP proteins with diverse recognition sequences, which may expand the current protease toolbox for specific removal of expression tags in protein chemistry or leader peptides in RiPP biosynthesis.

6.2 Results

6.2.1 Expression of LicP reveals a self-cleavage maturation process

The *licP* gene was amplified from genomic DNA of *B. licheniformis* ATCC 14580 (*25*) and cloned into an expression vector. A hexa-histidine tag was installed at its N-terminus, and the first 24 amino acids of LicP, which correspond to a secretion signal peptide, were omitted. Upon expression in *Escherichia coli* BL21 (DE3) and purification using immobilized metal affinity chromatography, two bands were observed by gel electrophoresis (**Figure 6.3a**). Analysis by MALDI-TOF MS demonstrated masses of 9,923 Da and 37,431 Da (**Figure 6.3b**). These molecular weights are in agreement with two fragments of LicP, an N-terminal portion His_6-LicP-25-100 with a calculated mass of 9,924 Da and a C-terminal portion LicP-101-433 with a calculated mass of 37,449 Da, suggesting that a cleavage event occurred during the expression of His_6-LicP-25-433 (hereafter referred to as wild type LicP).

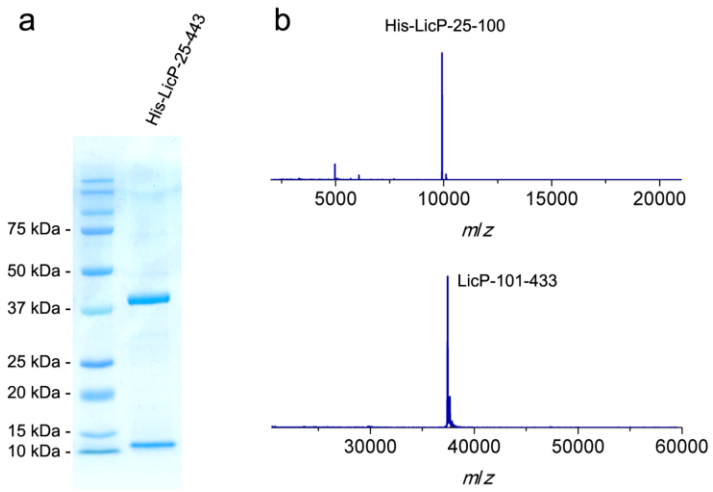

Figure 6.3 SDS-PAGE (a) and MALDI-TOF MS (b) analysis of His_6-LicP-25-433 expressed in *E. coli*. His_6-LicP-25-100, calculated M: 9,924, average mass; observed $M+H^+$: 9,923, average mass. LicP-101-433, calculated M: 37,431, average mass; observed $M+H^+$: 37,449, average mass.

Such proteolytic processing has been reported for several extracellular class I LanP proteases and was suggested to be autocatalytic (*12, 13*). To test whether this mechanism applied to the class II enzyme LicP, the predicted catalytic Ser376 was substituted by Ala. Unfortunately, His_6-LicP-25-433-S376A was expressed almost exclusively in the insoluble fraction. I also mutated His186 predicted to be involved in the catalytic triad, but His_6-LicP-25-433-H186A was again expressed insolubly. I eventually obtained a very small amount of soluble His_6-LicP-25-433-S376A, demonstrating that indeed the proteolytic cleavage after Glu100 was abolished (**Figure 6.4**). The observation that an inactivated LicP was expressed as the full length protein indicates the cleavage event is indeed catalyzed by LicP rather than *E. coli* proteases. These findings mirror those of a very recent report on a protease from a non-lanthipeptide producing organism that has sequence homology with the lanthipeptide protease EpiP, which employs an autocatalytic mechanism of cleavage between Lys87 and Thr88 (*26*). Using His_6-LicP-25-433 and its S376A mutant, I confirmed that such an autoproteolysis can take place intermolecularly, albeit slowly (**Figure 6.5**).

To obtain an active form of LicP with the pro-sequence covalently attached, Glu100was substituted with Ala. However, His_6-LicP-25-433-E100A was again expressed and purified as two fragments (**Figure 6.6a**). Surprisingly, the resulting fragments corresponded to a shifted cleavage site from residue Ala100 to Glu102 (**Figure 6.6b**). Further mutation of Glu102 to Ala abolished the production of soluble protein.

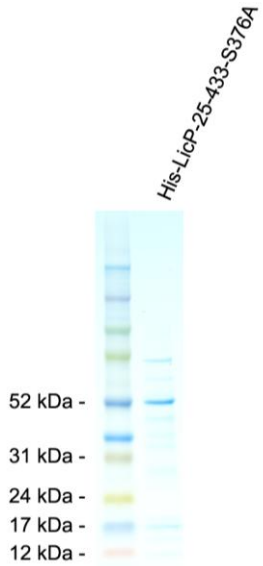

Figure 6.4 SDS-PAGE image of soluble His_6-LicP-25-433-S376A.

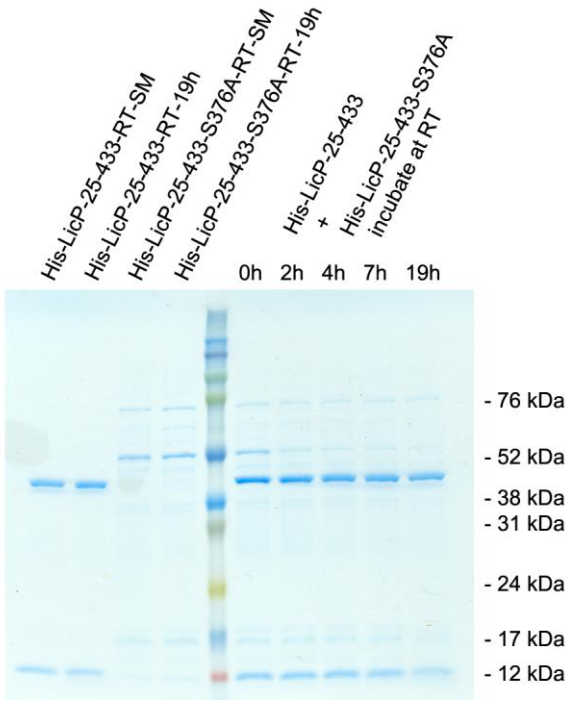

Figure 6.5 Determination if LicP cleavage is intra or intermolecular. To differentiate whether the self-cleavage of LicP occurred intramolecularly or intermolecularly, His_6-LicP-25-433 (consisting of a complex of His_6-LicP-25-100 and LicP-101-433) was incubated with His_6-LicP-25-433-S376A in a 1:1 ratio. The reaction was monitored by SDS-PAGE to determine whether wild type LicP catalyzes the proteolytic cleavage of His_6-LicP-25-433-S376A. When incubated alone, His_6-LicP-25-433 or His_6-LicP-25-433-S376A did not show any changes throughout the 19-hour incubation period, whereas the full length protein His_6-LicP-25-433-S376A was consumed gradually when incubated with His_6-LicP-25-433, suggesting that the observed cleavage of LicP can take place intermolecularly. Proteins were supplied with a final concentration of 0.1 mg/mL each.

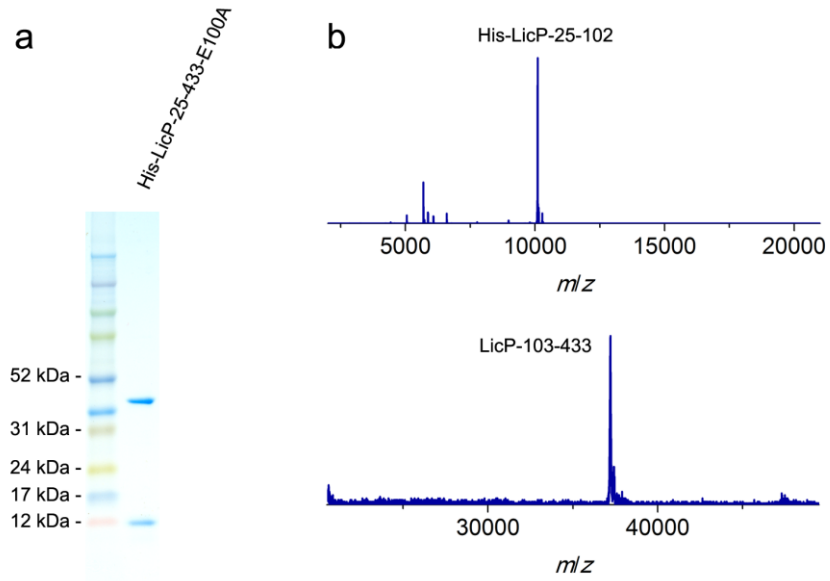

Figure 6.6 SDS-PAGE (a) and MALDI-TOF MS (b) analysis of His_6-LicP-25-433-E100A. His_6-LicP-25-102-E100A, calculated M: 10,096, average mass; observed $M+H^+$: 10,099, average mass. LicP-103-433, calculated M: 37,219, average mass; observed $M+H^+$: 37,207, average mass.

6.2.2 *In vitro* characterization of LicP

LicP has been suggested to trim off the 6-residue oligopeptide NDVNPE from LicA2' to generate mature Licβ (*18*). In this work, dehydrated and cyclized LicA2 was obtained by co-expressing LicA2 with its cognate lanthionine synthetase LicM2 in *E. coli* (**Figure 6.7a**) (*27*). Instead of using the membrane-bound protein LicT to produce LicA2', the commercial protease AspN was employed to generate a DVNPE-Licβ. Upon incubation with wild type LicP, the 5-residue oligopeptide was successfully removed (**Figure 6.7b**), confirming LicP's anticipated proteolytic activity against its substrate. When the full length modified LicA2 was incubated with LicP, the peptide was consumed, resulting in two fragments corresponding to the leader peptide and Licβ (**Figure 6.7c**). This observation suggests that LicP does not require prior proteolysis by LicT to produce Licβ. I next incubated linear LicA2 with LicP to test whether unmodified LicA2 was also a substrate (**Figure 6.8a**). Cleavage of the unmodified LicA2 peptide was observed (**Figure 6.8b**), indicating that post-translational modifications are not required for substrate recognition by LicP.

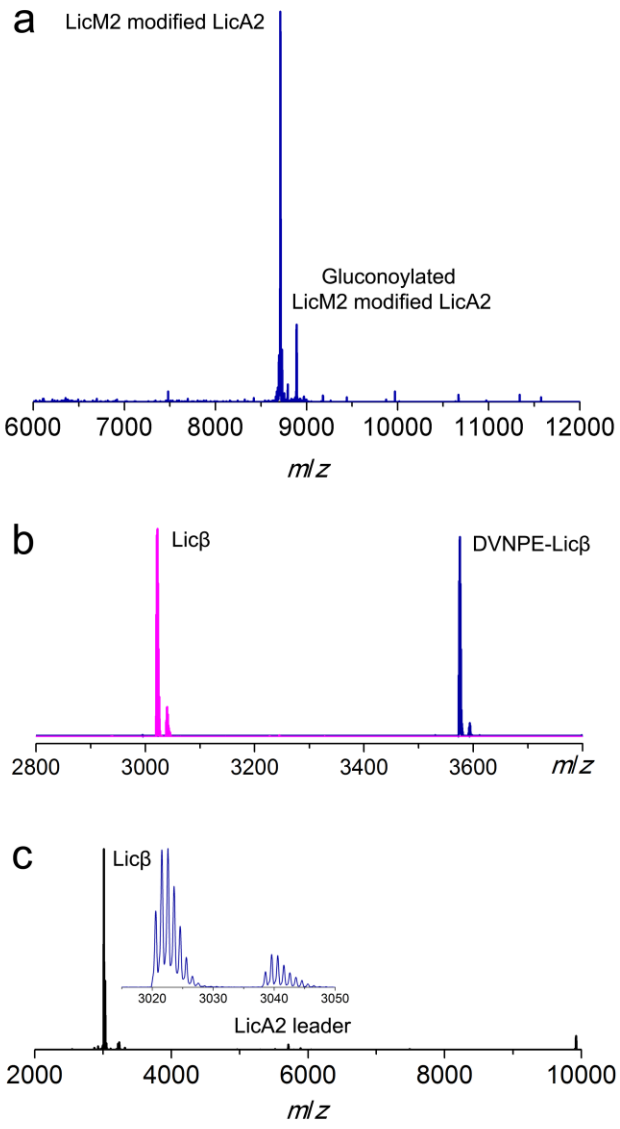

Figure 6.7 ***In vitro*** **activity of LicP against LicM2 modified LicA2. (a)** MALDI-TOF mass spectrum of LicM2-modified LicA2. LicM2 modified LicA2, calculated M-12H_2O: 8,714, average mass; observed M-12H_2O+H^+: 8,713, average mass. Gluconoylation at the N-terminus of LicA2 was introduced when expressing the peptide in *E. coli* BL21(DE3), resulting in a +178 Da peak in addition to the peak with the original mass (*28*). **(b)** MALDI-TOF mass spectra of DVNPE-Licβ peptide with (magenta) or without (blue) incubation with LicP. For DVNPE-Licβ, calculated M: 3572.6, monoisotopic mass; observed M+H^+: 3573.6, monoisotopic mass. For Licβ, calculated M: 3019.4, monoisotopic mass; observed M+H^+: 3020.5, monoisotopic mass. **(c)** MALDI-TOF mass spectrum for LicM2 modified LicA2 incubated with LicP. Licβ, calculated M: 3,019.4, monoisotopic mass; observed M+H^+: 3,020.6, monoisotopic mass. LicA2 leader peptide, calculated M: 5,711, average mass; observed M+H^+: 5,711, average mass.

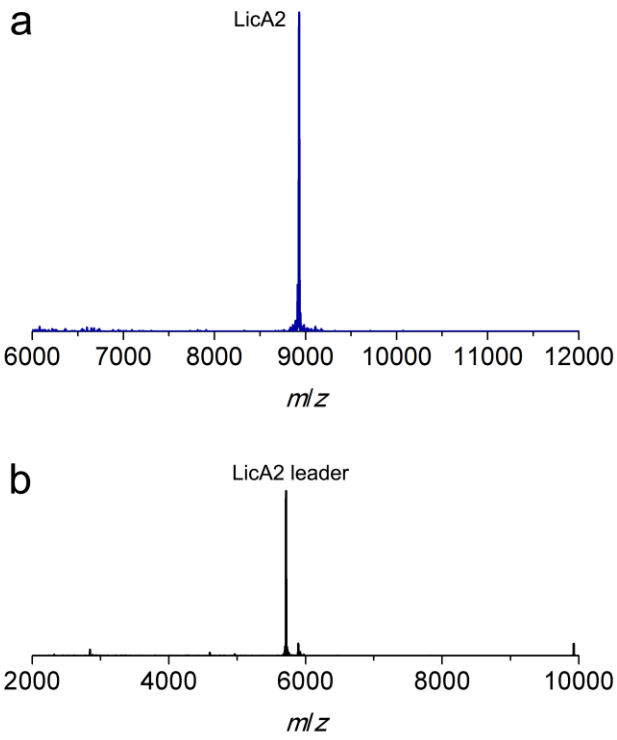

Figure 6.8 *In vitro* activity of LicP against linear LicA2. (a) MALDI-TOF mass spectrum of linear LicA2. LicA2, calculated M: 8,930, average mass; observed $M+H^+$: 8,929, average mass. **(b)** MALDI-TOF mass spectrum for linear LicA2 incubated with LicP. LicA2 leader peptide, calculated M: 5,711, average mass; observed $M+H^+$: 5,713, average mass. Unmodified LicA2 core peptide was not observed presumably due to poor ionization efficiency.

Subsequently, I investigated whether the enzyme displays a preference for modified or linear LicA2. Liquid chromatography-based kinetic analysis was hampered by the poor solubility of the LicA2 and Licβ peptides. Instead, a competitive MALDI-TOF MS assay was employed for a semi-quantitative time-dependent analysis, in which LicP was supplied to a mixture of modified and linear LicA2 and the production of leader peptides was monitored over time. In order to differentiate the otherwise identical leader peptides after proteolysis, a Pro to Gly mutation was introduced between the hexa-histidine tag and the precursor peptide in linear LicA2 (G-LicA2). The leader peptides obtained by complete proteolysis of equimolar amounts of modified and linear LicA2 exhibited comparable signal intensities when monitored by MALDI-TOF MS, confirming that the Pro to Gly mutation does not alter the ionization efficiency significantly (**Figure 6.9**). LicP was incubated with an 800-fold excess of modified and linear LicA2, and MALDI-TOF MS analysis illustrated complete consumption of modified LicA2 peptide within 10 min,

corresponding to a rate of at least 80 min^{-1}, whereas the cleavage of linear LicA2 only started after the modified LicA2 had been consumed and required more enzyme to be completed (**Figure 6.9**). Collectively, these observations indicate that although both are substrates for the enzyme, LicP strongly prefers modified LicA2 compared to the linear version.

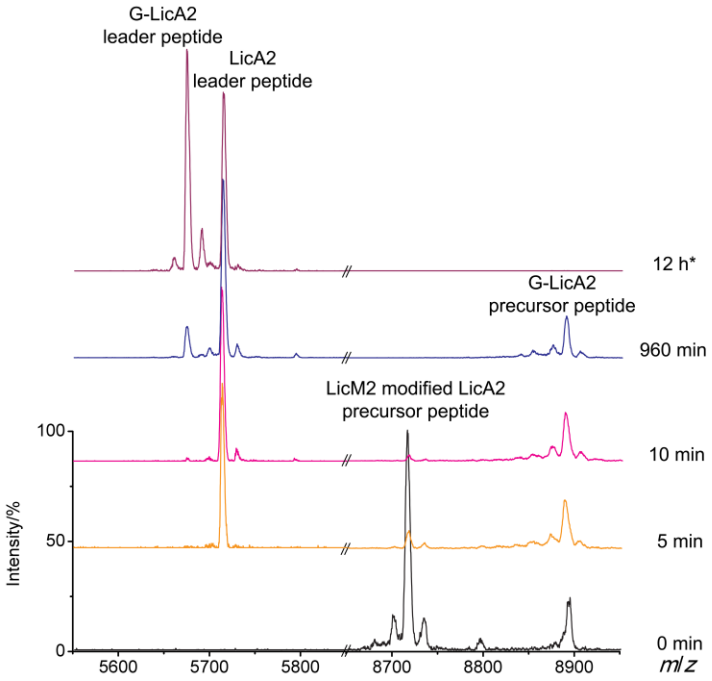

Figure 6.9 Time-dependent MALDI-TOF analysis of modified LicA2 and linear G-LicA2 peptides treated with LicP. For all traces, modified LicA2 and linear G-LicA2 were each supplied with a final concentration of 17 µM. For the purple trace, 2.1 µM of His_6-LicP-25-433 was added and the reaction was incubated at room temperature for 12 h (asterisk); for the other traces, 21 nM His_6-LicP-25-433 was employed. The intensity of the G-LicA2 leader peptide was set to 100% in the region of 5,500-5,900 Da, whereas the intensity of modified LicA2 precursor peptide was set to 100% in the region of 8,600-9,000 Da. Relative intensities are shown for comparison purposes.

6.2.3 LicP can serve as a sequence-specific traceless protease

The observation that LicP removes the oligopeptide NDVNPE and the entire leader peptide from modified or linear LicA2 suggests that it specifically recognizes the NDVNPE sequence but is rather tolerant of other regions of the peptides. This hypothesis was tested with other members of the lanthipeptide family, as site-specific removal of leader peptides is crucial for producing lanthipeptides *in vitro* and this step is often challenging as only a limited choice of proteases is available (*29*). ProcA1.7 and NisA, the precursor peptides of the lanthipeptides prochlorosin 1.7 and nisin, were mutated to substitute the last six residues of their leader peptides with the

NDVNPE sequence (**Figure 6.10**). ProcA1.7-NDVNPE and NisA-NDVNPE were incubated with LicP, resulting in successful removal of their leader peptides (**Figure 6.11**).

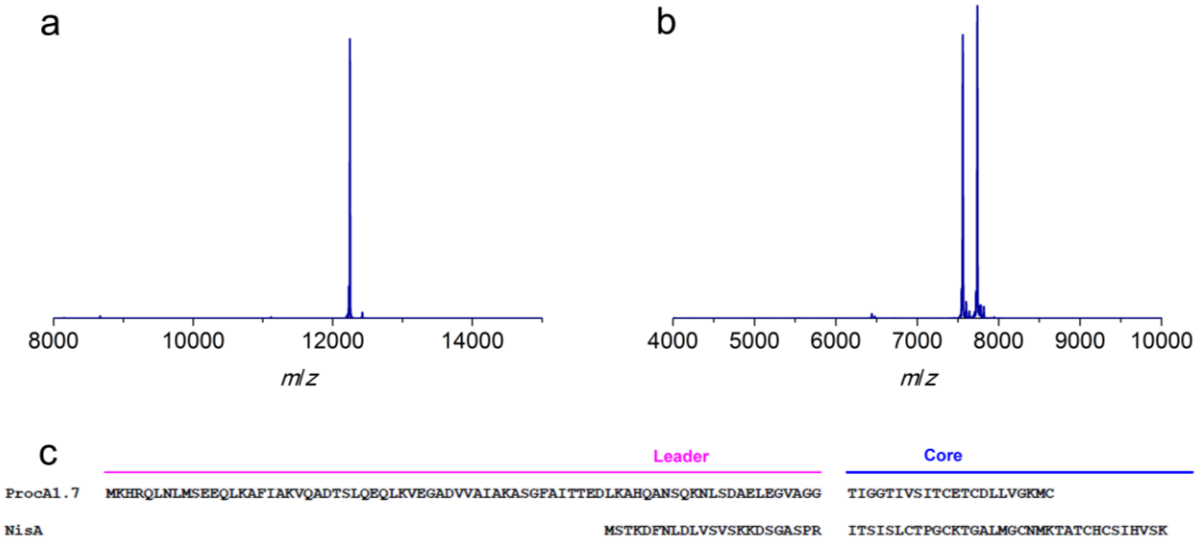

Figure 6.10 MALDI-TOF mass spectra for ProcA1.7-NDVNPE (a) and NisA-NDVNPE (b) and the primary sequences of NisA and ProcA1.7 (c). ProcA1.7-NDVNPE, calculated M: 12,244, average mass; observed M+H$^+$: 12,246, average mass. NisA-NDVNPE, calculated M: 7,557, average mass; observed M+H$^+$: 7,558, average mass. Gluconoylation at the N-terminus of NisA-NDVNPE was introduced when expressing the peptide in *E. coli* BL21(DE3), resulting in a +178 Da peak in addition to the peak with the original mass (*28*).

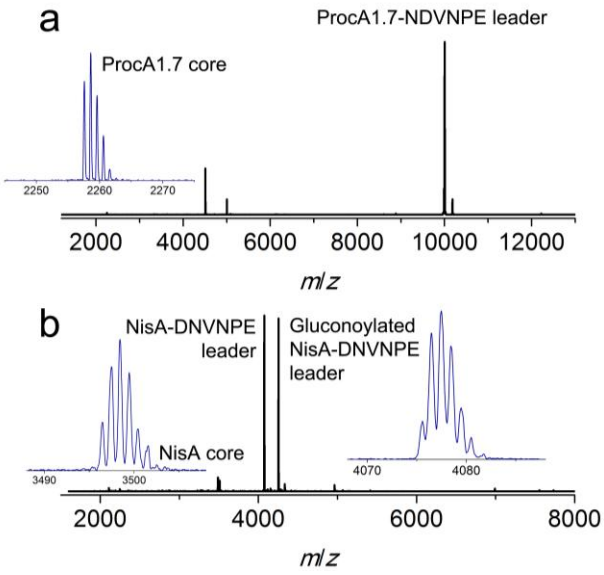

Figure 6.11 MALDI-TOF mass spectra for ProcA1.7-NDVNPE (a) and NisA-NDVNPE (b) incubated with LicP. ProcA1.7 core peptide, calculated M: 2,256.1, monoisotopic mass; observed M+H$^+$: 2,257.6, monoisotopic mass. ProcA1.7-NDVNPE leader peptide, calculated M: 10,004, average mass; observed M+H$^+$: 10,004, average mass. NisA core peptide, calculated M: 3,495.6, monoisotopic mass; observed M+H$^+$: 3,496.4, monoisotopic mass. NisA-NDVNPE leader peptide, calculated M: 4,074.9, monoisotopic mass; observed M+H$^+$: 4,075.5, monoisotopic mass.

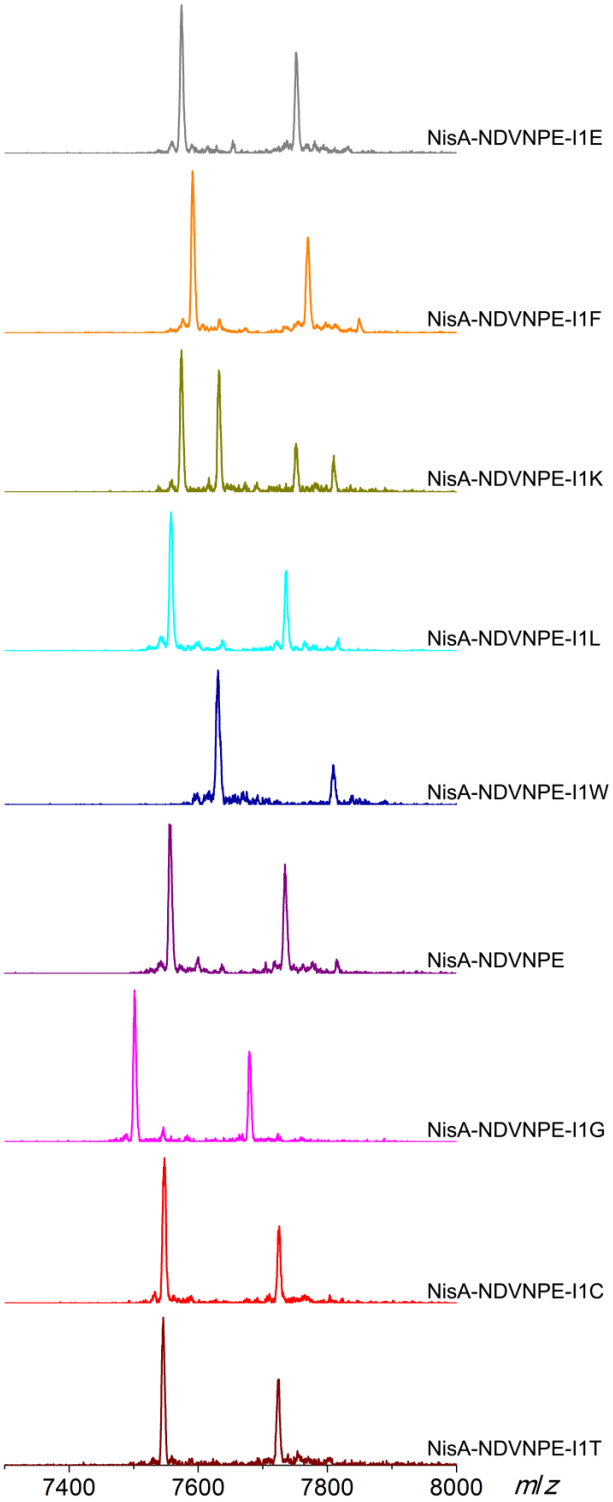

Figure 6.12 MALDI-TOF mass spectra for NisA-NDVNPE peptides with various P1' positions. Gluconoylation at the N-terminus of NisA-NDVNPE peptides was introduced when expressing these peptides in *E. coli* BL21(DE3), resulting in a +178 Da peak in addition to the peak of the original mass (*28*). Unknown modification of +58 Da was installed on NisA-NDVNPE-I1K, nevertheless, MALDI-TOF MS analysis suggested the mass peak corresponding to the correct mass as the major peak.

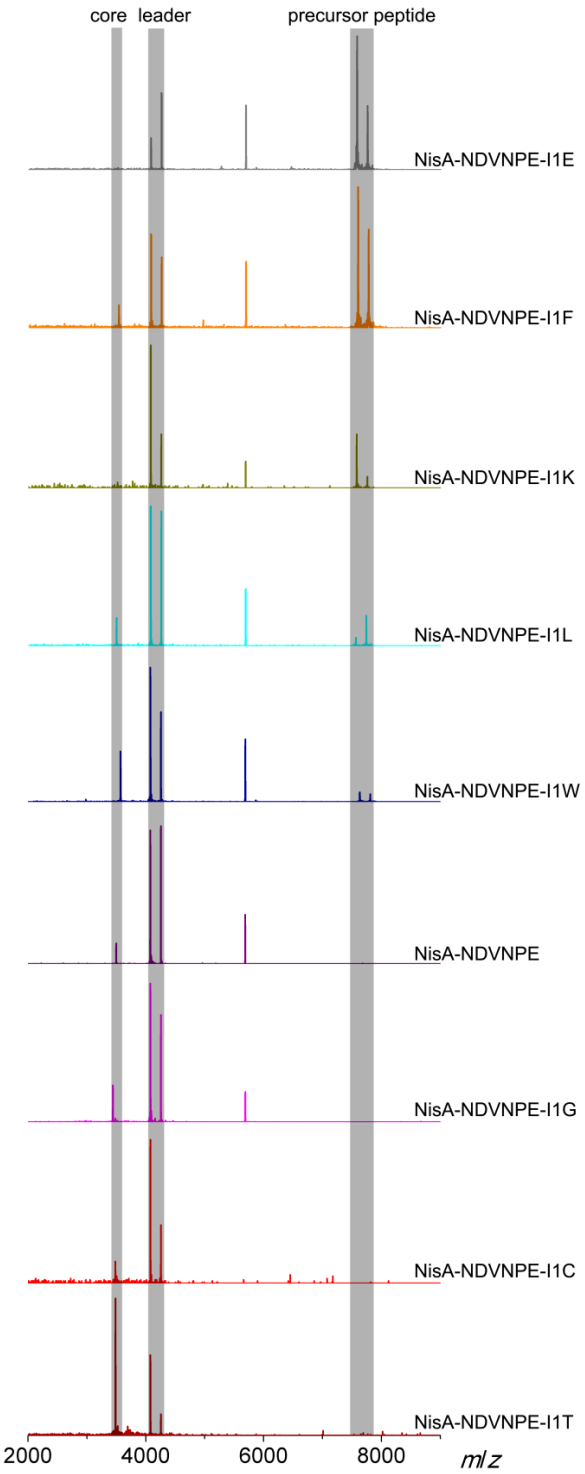

Figure 6.13 MALDI-TOF mass spectra for NisA-NDVNPE peptides with various P1' positions incubated with LicP. For all reactions, 65 µM of NisA variants were included. For NisA-NDVNPE-I1T and NisA-NDVNPE-I1C, LicP was supplied with a final concentration of 210 nM (enzyme:substrate = 310:1) and the reactions were incubated at room temperature for 20 hours. For other NisA variants, LicP was supplied with a final concentration of 2.1 µM (enzyme:substrate = 31:1) and the reactions were incubated at room temperature for 30 hours.

The experiments with linear and modified LicA as well as ProcA1.7-NDVNPE and NisA-NDVNPE demonstrated that LicP tolerates Dhb, Thr and Ile in the P1' position. To further evaluate its tolerance, the P1' position in NisA-NDVNPE was changed from Ile to eight other amino acids (Gly, Cys, Thr, Leu, Phe, Trp, Glu and Lys; **Figure 6.12**). All these mutants were accepted by LicP as in all cases removal of the NisA leader peptide was observed when substrate peptides were supplied in 30 to 300 fold excess over the enzyme (**Figure 6.13**). The highest proteolytic efficiency was obtained when the P1' position was occupied by Thr or Cys; traceless removal of tags in front of Cys is highly valuable for cysteine-based ligation chemistry (*30, 31*). The removal of the NisA leader peptide in front of a Gly or Ile residue was slightly less efficient, but complete consumption of precursor peptides was still observed. NisA-NDVNPE analogs with Trp, Leu, and Lys at the P1' position were also accepted by LicP, although some substrate still remained after 30 hours of incubation. Peptides with Glu or Phe in the P1' position turned out to be poor substrates. Collectively, these results show that LicP serves as a sequence-specific protease for non-native substrates and that its activity is highly portable with respect to the P1' position (**Table 6.1**).

Table 6.1 Peptides containing cleavage sites with different P and P' sequences that are accepted by LicP.

Substrate sequences	P and P' positions	[LicP]	[Substrate]	Incubated at room temperature	Reaction goes to completion
LicP-95-105	NTAVNE | TESVI	–	–	–	Y
LicP-E100A-97-107	AVNATE | SVISG	–	–	–	Y
Modified LicA2	NDVNPE | DhbDhbPADhb	21 nM	17 µM	10 min	Y
Linear LicA2	NDVNPE | TTPAT	1.1 µM	290 µM	6 h	Y
ProcA1.7-NDVNPE	NDVNPE | TIGGT	210 nM	180 µM	4 h	Y
NisA-NDVNPE	NDVNPE | ITSIS	2.1 µM	66 µM	30 h	Y
NisA-NDVNPE-I1T	NDVNPE | TTSIS	210 nM	66 µM	20 h	Y
NisA-NDVNPE-I1C	NDVNPE | CTSIS	210 nM	66 µM	20 h	Y
NisA-NDVNPE-I1G	NDVNPE | GTSIS	2.1 µM	67 µM	30 h	Y
NisA-NDVNPE-I1W	NDVNPE | WTSIS	2.1 µM	66 µM	30 h	N
NisA-NDVNPE-I1L	NDVNPE | LTSIS	2.1 µM	66 µM	30 h	N
NisA-NDVNPE-I1K	NDVNPE | KTSIS	2.1 µM	66 µM	30 h	N
NisA-NDVNPE-I1F	NDVNPE | FTSIS	2.1 µM	66 µM	30 h	N
NisA-NDVNPE-I1E	NDVNPE | ETSIS	2.1 µM	66 µM	30 h	N

6.2.4 X-ray structure of LicP

The 2.35 Å resolution structure of LicP reveals a peptidase_S8 family serine protease fold common among subtilisin-like enzymes (*32*) (**Figure 6.14a**). Although LicP was cleaved, a non-covalent complex of the resultant two fragments was observed in the crystal structure, consisting of an N-terminal prodomain and a C-terminal catalytic protease domain. Interestingly, in the LicP structure, the calcium-binding loop that is crucial for the stability of other subtilisin-like proteins is missing (**Figure 6.14b**) (*33*). Instead, Trp111 from the catalytic domain inserts into a hydrophobic pocket that situates at a similar place of the calcium-binding site in subtilisin BPN' (**Figure 6.14c**) and such an insertion may be a necessary step in LicP maturation. Perhaps the observed difficulty of obtaining soluble mutants of LicP that prevent proteolytic processing can be attributed to the inability of Trp111 to access this hydrophobic pocket.

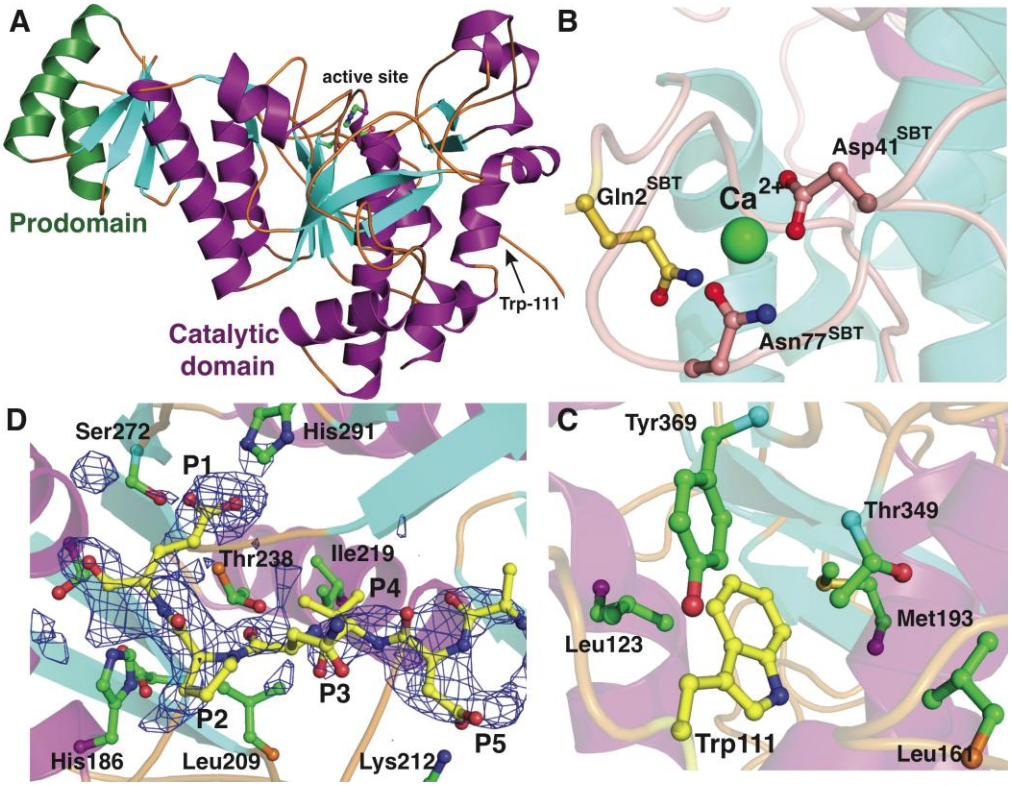

Figure 6.14 Crystal structure of LicP. (a) Cartoon representation of the overall structure of LicP with the prodomain colored in green and the catalytic domain colored in purple. **(b,c)** Comparison of the structures of LicP **(c)** and subtilisin BPN′ **(b)** in the vicinity of the calcium-binding region, which in LicP is replaced by a hydrophobic pocket into which Trp111 from the linker region binds. In panel b, the residues numbering refers to the subtilisin (SBT) catalytic domain, whereas in panel c and elsewhere LicP numbering is based on the full length protein. **(d)** Omit electron density map superimposed on a model of LicP bound to a peptide substrate. The density corresponds to residual features of the linker into which the substrate has been modeled. Figure made by Prof. Satish K. Nair.

In addition, the interaction of LicP with a NDVNPE peptide substrate was modeled using the spurious electron density observed in the vicinity of the LicP active site (**Figure 6.14d**). According to this model, the specificity for the P1 Glu is dictated by interactions with the side chains of Ser272 and His291, and weak interactions with the hydroxyl groups of Thr218 and Tyr241. Likewise, a pocket that consists of Ser272, His291, and Leu209 appears to dictate the P4 specificity for a small hydrophobic residue such as Val. Lastly, the Asp at the P5 position of the substrate is positioned to make ionic interactions with Lys212 of LicP. The paucity of additional interactions with the enzyme is consistent with the tolerance of LicP for the linker that connects the prodomain and the catalytic domain (sequence NTAVNE) and its substrate peptide (NDVNPE).

6.2.5 Class II LanP proteins: a pool of sequence-specific proteases

LanP genes are not often found in class II lanthipeptide gene clusters. Only four class II LanPs have been reported to date - LicP, CylA, CerP and CrnP, which have been suggested to remove six-residue sequences at the N-terminus of Licβ, cytolysin, cerecidins and carnolysin, respectively (*18, 34, 35*). Among them, the proteolytic activity, and hence the identity of the cleavage sites, has been confirmed for CylA (*17*), CerP (*34*) and LicP (this work). To identify additional class II LanP proteins and potentially identify additional recognition sequences that might be useful, I performed a search of the UniProtKB database with the LicP protease domain (LicP-101-433) as a query or the non-redundant protein sequence database with LicA2 as a query using the default Blast parameters for proteins provided by National Center for Biotechnology Information (NCBI) website. The first 250 hits were subjected to further analysis and quite a few are correlated to class II lanthipeptide biosynthesis with clustered LanM proteins. Nine representative class II lanthipeptide gene clusters with LanP genes were listed, all of which contain multiple genes for LanA substrates (**Figure 6.15** and **Table 6.2**). These LanP proteins share a minimal identity of 30% with LicP protease domain with E values lower than e^{-26}. The putative cleavage sites for these LanP enzymes were proposed to locate immediately C-terminal to the double Gly-type motif used by LanT enzymes and upstream of the Thr/Ser/Cys rich core peptides (**Figure 6.15**). Interestingly, all putative LanP recognition sequences consist of six residues with the exception of the A1-A3 peptides from *Bacillus cereus* FIR-35, which contain eight residues with two additional amino acids at the N-terminus. An Asp at the P5 position and a Val at P4 position are conserved among most clusters with the precursor peptides from *Bacillus cereus* VD045 being exceptions. Pro and

Ala are frequently found at the P2 position, whereas the P1 position is almost exclusively occupied by polar/charged residues, such as Asp, Glu, His or Arg, but the putative P1 position of the A1 peptide from *Bacillus thuringiensis* DB27 is occupied by Ala.

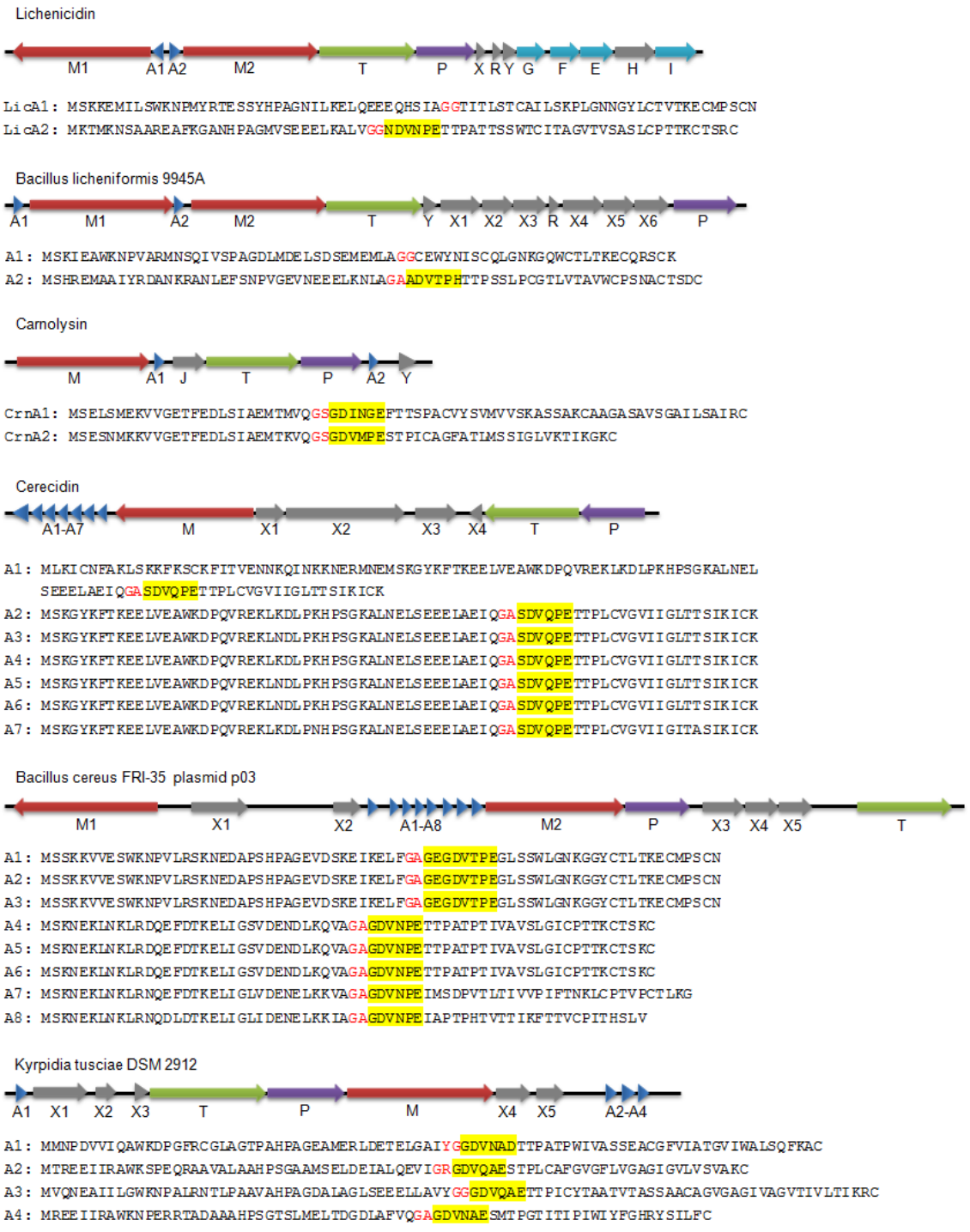

Figure 6.15 (cont. on next page)

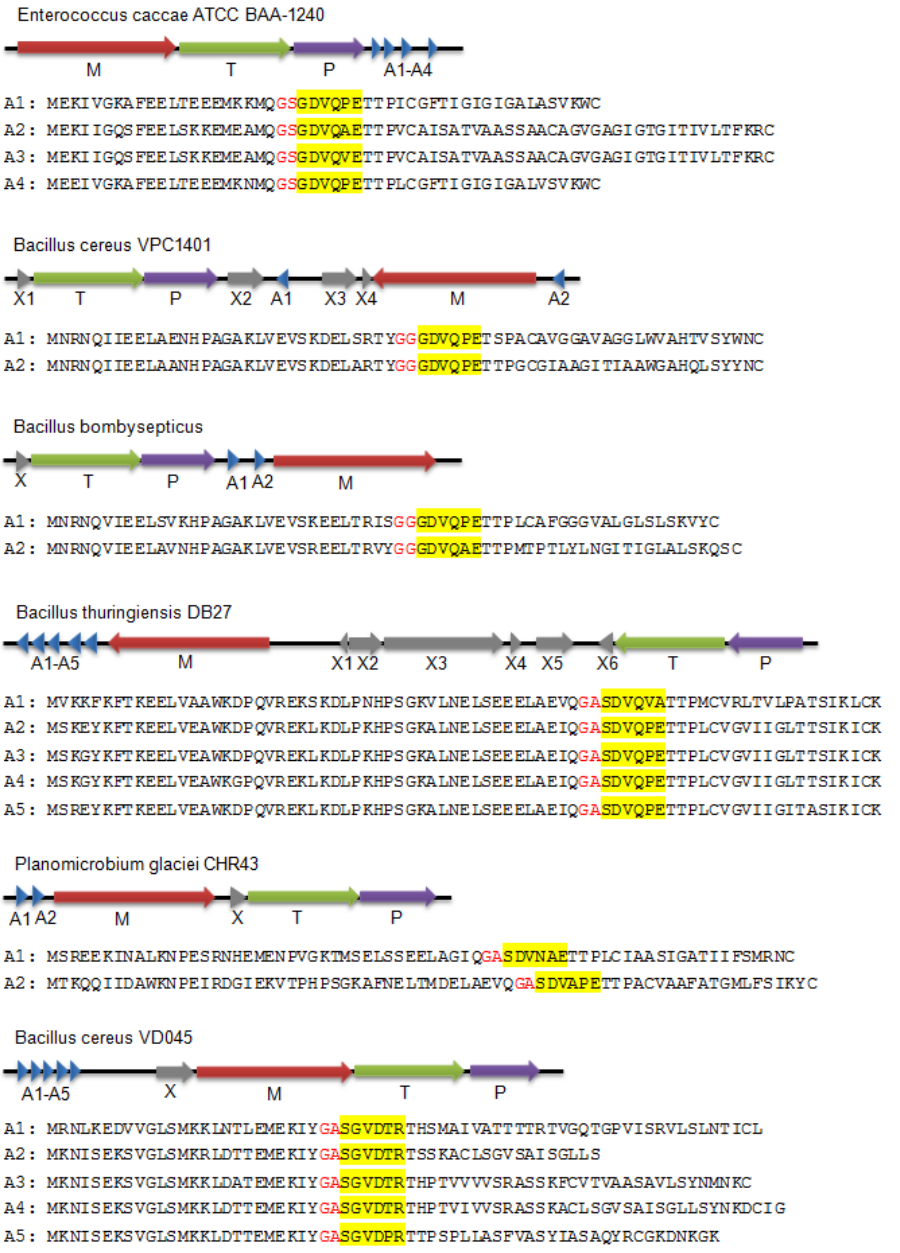

Figure 6.15 Thirteen class II lanthipeptide biosynthetic gene clusters containing LanP genes. Genes with unknown functions are indicated with X. Substrate LanA sequences are listed under the genetic pathways and the putative LanP recognition sequences are highlighted.

Table 6.2 Class II LanP proteins and their predicted secretion signal peptide sequences.
Secretion signal peptide sequences are predicted using an online tool PrediSi (*36*).

Name	Organism	Accession number	Signal peptide (residue)
LicP	*Bacillus licheniformis* ATCC 14580	AAU42937.1	1-24
CylA	*Eenterococcus faecalis*	AFJ74725.1	1-24
CerP	*Bacillus cereus* Q1	ACM15351.1	1-36
CrnP	*Carnobacterium maltaromaticum*	AHF21241.1	1-30
LanP	*Bacillus licheniformis* 9945A	AGN34600.1	1-37
LanP	*Bacillus cereus* FRI-35	AFQ13336.1	1-25
LanP	*Kyrpidia tusciae* DSM 2912	ADG07479.1	1-31
LanP	*Enterococcus caccae* ATCC BAA-1240	EOL44526.1	1-28
LanP	*Bacillus cereus* VPC1401	YP_004050051.1	1-31
LanP	Bacillus bombysepticus	AHX21587.1	1-31
LanP	*Bacillus thuringiensis* DB27	CDN38711.1	1-36
LanP	*Planomicrobium glaciei* CHR43	ETP67278.1	1-30
LanP	*Bacillus cereus* VD045	EJR29324.1	1-27

6.3 Discussion

In this chapter, I describe the heterologous expression and *in vitro* characterization of a class II LanP protease, a protein class that has not been experimentally characterized before. LicP was obtained from *E. coli* as two fragments that bound tightly to each other in a 1:1 ratio, and the observed cleavage is autocatalytic. Such an autocatalytic cleavage has been reported for two class I LanPs, NisP and EpiP, which lack the first 195 and 99 amino acids in their mature forms, respectively (*12*, *13*). The cleavage event was recently confirmed to be intramolecular for a homolog of EpiP (*26*). Here I show that for LicP, such cleavage can proceed intermolecularly, but because such cleavage requires hours of incubation, it is likely that the autoproteolysis also occurs intramolecularly under physiological conditions. Interestingly, the E100A mutant was still expressed as two fragments with a shift in the cleavage site by two residues to Glu102. This observation suggests that a Glu at the P1 position is essential for LicP activity. However, the cleavage at Glu102 is still surprising because residues 95-100 in LicP do not appear to be homologous to residues 97-102 in LicP-E100A except the cleavage site glutamate (**Table 6.1**), and because ProcA1.7-NDVNPE contains seven glutamates (**Figure 6.10**), and yet LicP only cleaved

after the engineered glutamate at position −1 under the conditions used. Presumably, the linker sequence for autocatalytic processing is precisely positioned to maintain a high effective concentration such that cleavage at alternative Glu sites is still relatively efficient.

In addition to its physiological role of trimming a six-residue oligopeptide from LicA2', LicP removed the entire leader peptides from modified and linear LicA2. A substrate consumption rate of >80 min^{-1} was observed for LicA2 with the post-translational modifications installed, indicating LicP is a highly efficient protease. NisP, the class I LanP responsible for the maturation of the lantibiotic nisin, *in vivo* only accepts modified NisA as its substrate but not the linear version (*37*, *38*). In accord with these findings, our observations suggest that LicP also favors modified substrate *in vitro*. Such substrate selectivity by LanP proteins is beneficial as it prevents undesired consumption of unmodified precursor peptides.

Although LicP favors modified LicA2 over linear LicA2, which indicates that post-translational modifications in the core peptide contribute to LicP's substrate recognition in addition to the NDVNPE sequence, my observations with substrate analogs demonstrate its application as a sequence-specific protease for traceless removal of leader peptides or potentially expression tags. Collectively, LicP accepted a range of residues at the P1' site (**Table 6.1**) such as glycine, small polar residues (Thr, Cys) and large aliphatic residues (Ile). Also substrates, but less efficient, were peptides with Leu, aromatic (Phe, Trp) and charged (Lys, Glu) residues at the P1' position. Additional favorable properties include its stability as LicP remained active after 12 weeks at 4 ^{o}C, and no obvious decrease of activity was observed after multiple rounds of freeze-thaw procedures.

The thermostability of subtilisin BPN′ and related proteases is enhanced significantly in the presence of calcium ions, which are necessary for maturation, and subsequent stabilization of a large loop in the catalytic domain (*39*). The calcium dependence constitutes a drawback for industrial utility of subtilisin BPN′. Much effort has been spent on engineering thermostable mutants of subtilisin that function in a calcium-independent manner (*40*, *41*). The structural and biochemical analysis of LicP reveals an elegant solution to this limitation, as maturation and subsequent stabilization of the enzyme is facilitated not by metal ions, but rather by the insertion of Trp111, liberated by cleavage of the linker between the prodomain and the catalytic domain, into a hydrophobic pocket located in the same vicinity as the calcium-binding site in subtilisin BPN′. A

recent structure of the class I lanthipeptide protease NisP also demonstrated loss of a calcium binding site, although unlike the structure of LicP, the prodomain was not present in the NisP structure (*22*).

Over the past several decades, the toolbox of useful proteases has been significantly enlarged. Several proteases with strict recognition sequences have been commercialized for biochemical or industrial applications, including factor Xa, enterokinase, and Tobacco Etch Virus (TEV) protease. Factor Xa and enterokinase exhibit trypsin-like activity and cleave after an Arg or Lys. TEV protease recognizes a larger motif and exhibits better reliability in terms of specificity, but TEV protease requires either a Gly or Ser at the P1' position for efficient cleavage. LicP is complementary in that it specifically cleaves after a Glu in the NDVNPE sequence, and is quite tolerant of various residues in the P1' position. I also identified nine new class II lanthipeptide gene clusters containing *lanP* genes, suggesting they are more widely distributed than previously expected (*11*). Although the putative recognition sequences of these newly identified LanPs show a certain level of homology, they also exhibit considerable diversities. Similar to other proteases, most of these LanPs are predicted to cleave after charged residues such as arginine, glutamate or aspartate, but a few appear to cleave after unusual P1 residues such as histidine or alanine that are rarely the site of cleavage for other proteases (**Figure 6.15**). Hence, this naturally occurring protease family may serve as a basis to construct a general protease pool for traceless tag removal purposes (*5, 42, 43*). For characterization of another class II LanP, CylA, see chapter 7.

6.4 Methods

General methods

Similar general methods and materials as what were described in chapter 2 were employed in this chapter unless specified otherwise. Solid phase extraction was performed with a Strata-X polymeric reverse phase column (Phenomenex). FPLC was carried out using an AKTA FPLC system (Amersham Pharmacia Biosystems). Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) was carried out on a Bruker Daltonics UltrafleXtreme MALDI-TOF/TOF mass spectrometer. The detection of peptides with low molecular weights (700-3,500 Da), peptides with medium molecular weights (3,500-20,000 Da) and proteins with high molecular weights (20,000-50,000 Da) was achieved by using different

instrument settings optimized for these mass ranges. C18 zip-tip pipet tips were obtained from Millipore to desalt samples for MS analysis. The lichenicidin producing strain, *Bacillus licheniformis* ATCC 14580, was obtained from American Type Culture Collection.

Extraction of genomic DNA from *Bacillus licheniformis* ATCC 14580

Bacillus licheniformis ATCC 14580 was cultured in LB medium at 37 ^{o}C aerobically for 12 h and the genomic DNA was extracted using an UltraClean microbial DNA isolation kit following the manufacturer's protocol.

Construction of pRSFDuet-1 derivatives for expression of LicP-25-433 and LicA2

licP and *licA2* genes were amplified from the genomic DNA of *Bacillus licheniformis* ATCC 14580 using appropriate primers and cloned into the multiple cloning site 1 (MCS1) of a pRSFDuet-1 vector to generate pRSFDuet-1/LicP-25-433 and pRSFDuet-1/LicA2 plasmids, respectively. Primer sequences are listed in **Table 6.3**.

Construction of pRSFDuet-1 derivatives for expression of ProcA1.7-NDVNPE and NisA-NDVNPE

Engineered peptide genes were generated by multi-step overlap extension PCR. First, the amplification of the 5' leader part was carried out by 30 cycles of denaturing (95 ℃ for 10 s), annealing (55 ℃ for 30 s), and extending (72 ℃ for 15 s) using forward primers for *procA1*.7 and *nisA* and appropriate leader peptide reverse primers containing the mutations (**Table 6.3**) to generate a forward megaprimer (FMP). In parallel, PCR reactions using forward primers and reverse primers for *procA1*.7 and *nisA* core peptides (**Table 6.3**) were performed to produce the 3' core fragments (termed reverse megaprimer, RMP). The 5' FMP fragment and 3' RMP fragment were purified by 2% agarose gel, combined in equimolar amounts and amplified using the same PCR conditions as above with *procA1*.7 and *nisA* primers. The resulting PCR products were purified, digested and then cloned into the MCS1 of a pRSFDuet-1 vector to generate pRSFDuet-1/ProcA1.7-NDVNPE and pRSFDuet-1/NisA-NDVNPE plasmids.

Construction of pRSFDuet-1 derivatives for expression of LicP-25-433-S376A, LicP-25-433-H186A, LicP-25-433-E100A, LicP-25-433-E100A-E102A, G-LicA2, NisA-NDVNPE-I1G, NisA-NDVNPE-I1T, NisA-NDVNPE-I1C, NisA-NDVNPE-I1L, NisA-NDVNPE-I1F, NisA-NDVNPE-I1W, NisA-NDVNPE-I1K and NisA-NDVNPE-I1E

The expression plasmids pRSFDuet-1/LicP-25-433-S376A, pRSFDuet-1/LicP-25-433-H186A, pRSFDuet-1/LicP-25-433-E100A, pRSFDuet-1/LicP-25-433-E100A-E102A, pRSFDuet-1/G-LicA2, pRSFDuet-1/NisA-NDVNPE-I1G, pRSFDuet-1/NisA-NDVNPE-I1T, pRSFDuet-1/NisA-NDVNPE-I1C, pRSFDuet-1/NisA-NDVNPE-I1L, pRSFDuet-1/NisA-NDVNPE-I1F, pRSFDuet-1/NisA-NDVNPE-I1W, pRSFDuet-1/NisA-NDVNPE-I1K and pRSFDuet-1/NisA-NDVNPE-I1E were generated using QuikChange methodology based on pRSFDuet-1/LicP-25-433, pRSFDuet-1/LicA2 and pRSFDuet-1/NisA-NDVNPE as templates. Primer sequences are listed in **Table 6.3**.

Construction of pRSFDuet-1 derivatives for co-expression of LicM2 with LicA2

LicM2 was amplified from the genomic DNA of *Bacillus licheniformis* ATCC 14580 using appropriate primers and cloned into the MCS2 of a pRSFDuet-1 vector to generate pRSFDuet-1/LicM2-2. The expression plasmid pRSFDuet-1/LicA2/LicM2-2 was constructed by inserting the *licA2* gene into the MCS1 of the pRSFDuet-1/LicM2-2 plasmid. Primer sequences are listed in **Table 6.3**.

Table 6.3 Primer sequences for cloning of *licP-25-433, licM2, licA2, licP-25-433-S376A, licP-25-433-H186A, licP-25-433-E100A, licP-25-433-E100A-E102A, G-licA2, procA1.7-NDVNPE, nisA-NDVNPE, nisA-NDVNPE-I1G, nisA-NDVNPE-I1T, nisA-NDVNPE-I1C, nisA-NDVNPE-I1L, nisA-NDVNPE-I1F, nisA-NDVNPE-I1W, nisA-NDVNPE-I1K and nisA-NDVNPE-I1E*.

Primer Name	Primer Sequence (5'-3')
LicP_25_BamHI_FP	AAAAA GGATCCG AAAGAACAAGCAGGAGAACAG
LicP_NotI_RP	AAAAA GCGGCCGC TCACTCCTTG TTCATCATTT T
LicA2_BamHI_FP	AAAAA GGATCCG ATGAAAACAA TGAAAAATTC A
LicA2_NotI_RP	AAAAA GCGGCCGC CTAGCATCGG CTTGTACACT T
LicM2_NdeI_FP	AAAAA CATATG GTTTTCT TCGCCAAAGG GATG
LicM2_KpnI_RP	AAAAA GGTACC TCACCTGCCC GTCGGAATAT C
G-LicA2_ _QC_FP	CCAGGAT GGT ATGAAAAC AATGAAA AATTCAGCTGCCCGT

Table 6.3 (cont.)

G-LicA2_QC_RP	GTTTTCAT ACC ATCCTGG CT GTGGTGATGA TGGTGATGG
ProcA1.7_EcoRI_FP	GGT GCG AGG AAT TCG ATG AAG CAT AGA CAA CTA AAT CTG
ProcA1.7_NotI_RP	ATA ATA TCG CGG CCG CTC AGC ACA TTT TCC C
NisA_BamHI_FP	CTA GAT GGA TCC GAT GAG TAC AAA AGA TTT TAA CTT GG
NisA_HindIII_RP	CTA GAA GCT TTT ATT TGC TTA CGT GAA TAC TAC AAT G
LicP-S376A_QC_FP	GGAACA GCA TTGGCC GCCCCG CAGGTAGCT
LicP-S376A_QC_RP	GG CCAA TGC TGT TCC GTATGAG AGGGAATATC CCTTTGGGAT
LicP-H186A_QC_FP	ACA GGA GCC GGAAC ACAA ACAGCCGGGATGATCAATATC
LicP-H186A_QC_RP	G TTCC GGC TCC TGT CGGATCT CCGGATACAG GC
LicP-E100A_QC_FP	CAGTAAAC GCA ACGGAATC A GTCATCAGCGGTTCGCC
LicP-E100A_QC_RP	GATTCCGTTGCGTTTACTG CTGTATT TGCAATCGGC TTTTCAATGAC
LicP-E100A-E102A_QC_FP	GCAACG GCA TCAGTC ATCAGCGGTTCGCCTG
LicP-E100A-E102A _QC_RP	GACTGA TGC CGTTGC GTTTA CTGCTGTATT TGCAATCGGC
ProcA1.7_core_FP	ACCATTGGGGGA ACCATTGTG
ProcA1.7- NDVNPE _RP	GGTTCC CCCAATGGT TTCAGGATTGACGTCATT CAG CTCAGCATCA GACAGGT
NisA_core _FP	ATTACAAGTATTTCGCTATGT
NisA- NDVNPE _RP	CGAAATACTT GTAAT TTCAGGATTGACGTCATT ATCTTTC TTCGAAACAG ATA
NisA-NDVNPE-I1G_QC_FP	CAATCCTGAA GGT ACAAGTATTTC GCTATGTACACC CGGTTGTAAAAC
NisA-NDVNPE-I1G_QC_RP	GAAATACTTGT ACC TTCAGGATTG ACGTCATTATCTTTCTTCGAAACAGATACC
NisA-NDVNPE-I1C_QC_FP	CAATCCTGAA TGT ACAAGTATTTC GCTATGTACACC CGGTTGTAAAAC
NisA-NDVNPE-I1C_QC_RP	GAAAT ACTTGT ACA TTCAGGATTG ACGTCATTATCTTTCTTCGAAACAGATACC
NisA-NDVNPE-I1T_QC_FP	CAATCCTGAA ACC ACAAGTATTTC GCTATGTACACC CGGTTGTAAAAC AG
NisA-NDVNPE-I1T_QC_RP	GAAATACTTGT GGT TTCAGGATTG ACGTCATTATCTTTCTTCGAAACAGATACCA
NisA-NDVNPE-I1L_QC_FP	CAATCCTGAA CTT ACAAGTATTTC GCTATGTACACC CGGTTGTAAAAC
NisA-NDVNPE-I1L_QC_RP	GAAATACTTGT AAG TTCAGGATTG ACGTCATTATCTTTCTTCGAAACAGATACC
NisA-NDVNPE-I1F_QC_FP	CAATCCTGAA TTT ACAAGTATTTC GCTATGTACACC CGGTTGTAAAAC

Table 6.3 (cont.)

NisA-NDVNPE-I1F_QC_RP	GAAATACTTGT AAA TTCAGGATTG ACGTCATTATCTTTCTTCGAAACAGATACC
NisA-NDVNPE-I1W_QC_FP	CAATCCTGAA TGG ACAAGTATTTC GCTATGTACACC CGGTTGTAAAAC
NisA-NDVNPE-I1W_QC_RP	GAAATACTTGT CCA TTCAGGATTG ACGTCATTATCTTTCTTCGAAACAGATACC
NisA-NDVNPE-I1K_QC_FP	CAATCCTGAA AAA ACAAGTATTTC GCTATGTACACC CGGTTGTAAAAC
NisA-NDVNPE-I1K_QC_RP	GAAATACTTGT TTT TTCAGGATTG ACGTCATTATCTTTCTTCGAAACAGATACC
NisA-NDVNPE-I1E_QC_FP	CAATCCTGAA GAA ACAAGTATTTC GCTATGTACACC CGGTTGTAAAAC
NisA-NDVNPE-I1E_QC_RP	GAAATACTTGT TTC TTCAGGATTG ACGTCATTATCTTTCTTCGAAACAGATACC

Expression and purification of LicP and LicP mutant proteins

E. coli BL21 (DE3) cells were transformed with one of the following plasmids: pRSFDuet-1/LicP-25-433, pRSFDuet-1/LicP-25-433-S376A, pRSFDuet-1/LicP-25-433-H186A, pRSFDuet-1/LicP-25-433-E100A or pRSFDuet-1/LicP-25-433-E100A-E102A, and plated on an LB plate containing 50 mg/L kanamycin. A single colony was picked and grown in 20 mL of LB with kanamycin at 37 ^{o}C for 12 h and the resulting culture was inoculated into 2 L of LB. Cells were cultured at 37 ^{o}C until the OD at 600 nm reached 0.5, cooled and IPTG was added to a final concentration of 0.1 mM. The cells were cultured at 18 ^{o}C for another 10 h before harvesting. The cell pellet was resuspended on ice in LanP buffer (20 mM HEPES, 1 M NaCl, pH 7.5 at 25 ^{o}C) and lysed by homogenization. The lysed sample was centrifuged at 23,700 ×*g* for 30 min and the pellet was discarded. The supernatant was passed through 0.45-μm syringe filters and the protein was purified by immobilized metal affinity chromatography (IMAC) loaded with nickel as previously described (*44*). The proteins were generally eluted from the column at an imidazole concentration between 150 mM and 300 mM and the buffer was exchanged using a GE PD-10 desalting column or a gel-filtration column pre-equilibrated with LanP buffer. Protein concentration was quantified by its absorbance at 280 nm. The extinction coefficient for His_6-LicP-25-433 was calculated as 46,300 M^{-1} cm^{-1}. His_6-LicP-25-433-S376A was expressed dominantly in the inclusion body. Soluble protein was obtained by combining fractions eluted from the nickel column with the desired protein and concentrating to a small volume. No gel filtration was performed for the

mutant protein. Yield was determined to be 50 μg for 1 L culture. Aliquoted protein solutions were flash-frozen and kept at −80 ^{o}C until further usage.

Expression and purification of modified His$_6$-LicA2

Modified LicA2 was obtained using a similar procedure described previously using the corresponding co-expression vector (*45*, *46*).

Expression and purification of unmodified His$_6$-LicA2, His$_6$-G-LicA2, His$_6$-ProcA1.7-NDVNPE, His$_6$-NisA-NDVNPE, His$_6$-NisA-NDVNPE-I1G, His$_6$-NisA-NDVNPE-I1T, His$_6$-NisA-NDVNPE-I1C, His$_6$-NisA-NDVNPE-I1L, His$_6$-NisA-NDVNPE-I1F, His$_6$-NisA-NDVNPE-I1W, His$_6$-NisA-NDVNPE-I1K and His$_6$-NisA-NDVNPE-I1E

E. coli BL21 (DE3) cells were transformed with one of the following plasmids: pRSFDuet-1/LicA2, pRSFDuet-1/G-LicA2, pRSFDuet-1/ProcA1.7-NDVNPE, pRSFDuet-1/NisA-NDVNPE, pRSFDuet-1/NisA-NDVNPE-I1G, pRSFDuet-1/NisA-NDVNPE-I1T, pRSFDuet-1/NisA-NDVNPE-I1C, pRSFDuet-1/NisA-NDVNPE-I1L, pRSFDuet-1/NisA-NDVNPE-I1F, pRSFDuet-1/NisA-NDVNPE-I1W, pRSFDuet-1/NisA-NDVNPE-I1K or pRSFDuet-1/NisA-NDVNPE-I1E. Then the cells were plated on an LB plate containing 50 mg/L kanamycin. A single colony was picked and grown in 10 mL of LB with kanamycin at 37 ^{o}C for 12 h and the resulting culture was inoculated into 1 L of LB. Cells were cultured at 37 ^{o}C until the OD at 600 nm reached 0.5 and IPTG was added to a final concentration of 0.2 mM. The cells continued to be cultured at 37 ^{o}C for another 3 h before harvesting. The cell pellet was resuspended at room temperature in LanA start buffer (20 mM NaH_2PO_4, pH 7.5 at 25 ^{o}C, 500 mM NaCl, 0.5 mM imidazole, 20% glycerol) and lysed by sonication. The sample was centrifuged at 23,700 ×*g* for 30 min and the supernatant was discarded. The pellet was then resuspended in LanA buffer 1 (6 M guanidine hydrochloride, 20 mM NaH_2PO_4, pH 7.5 at 25 ^{o}C, 500 mM NaCl, 0.5 mM imidazole) and sonicated again. The insoluble portion was removed by centrifugation at 23,700 ×*g* for 30 min and the soluble portion was passed through 0.45-μm syringe filters. His-tagged peptides were purified by IMAC as previously described (*44*). The eluted fractions were desalted using reverse

phase HPLC or a Strata X polymeric reverse phase SPE column. The desalted peptides were lyophilized and stored at −20 °C for future use.

Intermolecular cleavage of His_6-LicP-25-433-S376A by His_6-LicP-25-433

His_6-LicP-25-433-S376A and His_6-LicP-25-433 proteins were both diluted with LanP buffer to a final concentration of 0.2 mg/mL. Parallel reactions were set up for His_6-LicP-25-433 only with a final protein concentration of 0.1 mg/mL in LanP buffer, His_6-LicP-25-433-S376A only with a final protein concentration of 0.1 mg/mL in LanP buffer, and His_6-LicP-25-433-S376A and His_6-LicP-25-433 together with a final protein concentration of 0.1 mg/mL each. The three reactions were allowed to proceed at room temperature for 0, 2, 4, 7 and 19 h before being stopped by addition of SDS loading buffer and boiling at 95 °C for 10 min.

Sequential proteolytic cleavage of modified LicA2

HPLC-purified LicM2-modified LicA2 was dissolved in H_2O to a final concentration of 3 mg/mL (340 μM). To a 17 μL solution of peptide (final peptide concentration 290 μM), 2 μL of 500 mM HEPES buffer (pH 7.5) was added followed by 1 μL of 0.5 mg/mL AspN. The reaction mixture was kept at room temperature for 12 h before 0.5 μL of 0.1 mg/mL LicP (final protein concentration 50 nM) was added. The reaction was then incubated at room temperature for one more hour. MALDI-TOF MS analysis was performed after each step.

Removal of leader peptides of modified or linear LicA2

Modified or linear LicA2 peptides were dissolved in H_2O to make a 3 mg/mL solution (340 μM). To a 17 μL solution of peptide (final peptide concentration 290 μM), 2 μL of 500 mM HEPES buffer (pH 7.5) was added followed by 1 μL of 1 mg/mL LicP (final protein concentration 1.1 μM). The reaction was incubated at room temperature for 6 h followed by MS analysis.

Competition assay of LicP activity with modified and linear LicA2

To a reaction vessel with 70 μL deionized H_2O, 5 μL each of 3 mg/mL modified LicA2 and linear G-LicA2 peptides were added (final peptide concentration 17 μM each) followed by 10 μL of 500 mM HEPES buffer (pH 7.5). Then, 10 μL of 0.01 mg/mL LicP was supplied (final protein

concentration 21 nM) and the reaction was incubated at room temperature before being stopped by addition of formic acid to a final concentration of 1% at different time points. To observe the complete consumption of both peptides, the same concentration and combination of substrate peptides were employed, to which 10 μL of 1 mg/mL LicP was added (final protein concentration 2.1 μM). The reaction mixture was kept at room temperature for 12 h before being stopped with 1% formic acid for MS analysis.

Proteolytic cleavage of the leader peptides of engineered peptides

ProcA1.7-NDVNPE was dissolved in H_2O to a final concentration of 3 mg/mL (250 μM), whereas for NisA-NDVNPE and its mutant peptides, a 10 mg/mL peptide solution was made (1.3 mM). For ProcA1.7-NDVNPE, 15 μL of peptide solution (final peptide concentration 190 μM) was pre-mixed with 1 μL of 50 mM DTT and 2 μL of 500 mM HEPES buffer (pH 7.5), to which 2 μL of 0.1 mg/mL LicP (final protein concentration 210 nM) was added. The reaction was incubated at room temperature for 4 h before analysis. For NisA-NDVNPE-I1T and NisA-NDVNPE-I1C, 1 μL of peptide (final peptide concentration 65 μM) was pre-mixed with 1 μL of 50 mM DTT and 2 μL of 500 mM HEPES buffer (pH 7.5) in 14 μL H_2O, to which 2 μL of 0.1 mg/mL LicP (final protein concentration 210 nM) was added. The reaction was incubated at room temperature for 20 h before analysis. For NisA-NDVNPE and other NisA mutant peptides, 1 mg/mL LicP (final protein concentration 2.1 μM) was employed instead of 0.1 mg/mL and the reaction was kept at room temperature for 30 h before being stopped for MS analysis.

Crystallization of LicP

This part was performed by Dr. Shi-Hui Dong and Prof. Satish K. Nair.

6.5 References

1. Rao, M. B., Tanksale, A. M., Ghatge, M. S., and Deshpande, V. V. (1998) Molecular and biotechnological aspects of microbial proteases, *Microbiol. Mol. Biol. Rev. 62*, 597-635.
2. Pogson, M., Georgiou, G., and Iverson, B. L. (2009) Engineering next generation proteases, *Curr. Opin. Biotechnol. 20*, 390-397.
3. Li, Q., Yi, L., Marek, P., and Iverson, B. L. (2013) Commercial proteases: present and future, *FEBS Lett. 587*, 1155-1163.
4. Craik, C. S., Page, M. J., and Madison, E. L. (2011) Proteases as therapeutics, *Biochem. J. 435*, 1-16.
5. Yi, L., Gebhard, M. C., Li, Q., Taft, J. M., Georgiou, G., and Iverson, B. L. (2013) Engineering of TEV protease variants by yeast ER sequestration screening (YESS) of combinatorial libraries, *Proc. Natl Acad. Sci. U.S.A. 110*, 7229-7234.
6. Chen, K., and Arnold, F. H. (1993) Tuning the activity of an enzyme for unusual environments: sequential random mutagenesis of subtilisin E for catalysis in dimethylformamide, *Proc. Natl Acad. Sci. U.S.A. 90*, 5618-5622.
7. Berg, D. T., Gerlitz, B., Shang, J., Smith, T., Santa, P., Richardson, M. A., Kurz, K. D., Grinnell, B. W., Mace, K., and Jones, B. E. (2003) Engineering the proteolytic specificity of activated protein C improves its pharmacological properties, *Proc. Natl Acad. Sci. U.S.A. 100*, 4423-4428.
8. Hedstrom, L., Perona, J. J., and Rutter, W. J. (1994) Converting trypsin to chymotrypsin: residue 172 is a substrate specificity determinant, *Biochemistry 33*, 8757-8763.
9. Rubingh, D. N. (1997) Protein engineering from a bioindustrial point of view, *Curr. Opin. Biotechnol. 8*, 417-422.
10. Arnison, P. G., Bibb, M. J., Bierbaum, G., Bowers, A. A., Bugni, T. S., Bulaj, G., Camarero, J. A., Campopiano, D. J., Challis, G. L., Clardy, J., Cotter, P. D., Craik, D. J., Dawson, M., Dittmann, E., Donadio, S., Dorrestein, P. C., Entian, K. D., Fischbach, M. A., Garavelli, J. S., Goransson, U., Gruber, C. W., Haft, D. H., Hemscheidt, T. K., Hertweck, C., Hill, C., Horswill, A. R., Jaspars, M., Kelly, W. L., Klinman, J. P., Kuipers, O. P., Link, A. J., Liu, W., Marahiel, M. A., Mitchell, D. A., Moll, G. N., Moore, B. S., Muller, R., Nair, S. K., Nes, I. F., Norris, G. E., Olivera, B. M., Onaka, H., Patchett, M. L., Piel, J., Reaney, M. J., Rebuffat, S., Ross, R. P., Sahl, H. G., Schmidt, E. W., Selsted, M. E., Severinov, K., Shen, B., Sivonen, K., Smith, L., Stein, T., Sussmuth, R. D., Tagg, J. R., Tang, G. L., Truman, A. W., Vederas, J. C., Walsh, C. T., Walton, J. D., Wenzel, S. C., Willey, J. M., and van der Donk, W. A. (2013) Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature, *Nat. Prod. Rep. 30*, 108-160.
11. Knerr, P. J., and van der Donk, W. A. (2012) Discovery, biosynthesis, and engineering of lantipeptides, *Annu. Rev. Biochem. 81*, 479-505.

12. van der Meer, J. R., Polman, J., Beerthuyzen, M. M., Siezen, R. J., Kuipers, O. P., and De Vos, W. M. (1993) Characterization of the *Lactococcus lactis* nisin A operon genes nisP, encoding a subtilisin-like serine protease involved in precursor processing, and nisR, encoding a regulatory protein involved in nisin biosynthesis, *J. Bacteriol. 175*, 2578-2588.
13. Geissler, S., Götz, F., and Kupke, T. (1996) Serine protease EpiP from *Staphylococcus epidermidis* catalyzes the processing of the epidermin precursor peptide, *J. Bacteriol. 178*, 284-288.
14. Velásquez, J. E., Zhang, X., and van der Donk, W. A. (2011) Biosynthesis of the antimicrobial peptide epilancin 15X and its unusual N-terminal lactate moiety, *Chem. Biol. 18*, 857-867.
15. Nishie, M., Shioya, K., Nagao, J., Jikuya, H., and Sonomoto, K. (2009) ATP-dependent leader peptide cleavage by NukT, a bifunctional ABC transporter, during lantibiotic biosynthesis, *J. Biosci. Bioeng. 108*, 460-464.
16. Furgerson Ihnken, L. A., Chatterjee, C., and van der Donk, W. A. (2008) *In vitro* reconstitution and substrate specificity of a lantibiotic protease, *Biochemistry 47*, 7352-7363.
17. Booth, M. C., Bogie, C. P., Sahl, H. G., Siezen, R. J., Hatter, K. L., and Gilmore, M. S. (1996) Structural analysis and proteolytic activation of *Enterococcus faecalis* cytolysin, a novel lantibiotic, *Mol. Microbiol. 21*, 1175-1184.
18. Caetano, T., Krawczyk, J. M., Mosker, E., Süssmuth, R. D., and Mendo, S. (2011) Heterologous expression, biosynthesis, and mutagenesis of type II lantibiotics from *Bacillus licheniformis* in *Escherichia coli*, *Chem. & Biol. 18*, 90-100.
19. Völler, G. H., Krawczyk, B., Ensle, P., and Süssmuth, R. D. (2013) Involvement and unusual substrate specificity of a prolyl oligopeptidase in class III lanthipeptide maturation, *J. Am. Chem. Soc. 135*, 7426-7429.
20. Geissler, S., Götz, F., and Kupke, T. (1996) Serine protease EpiP from *Staphylococcus epidermidis* catalyzes the processing of the epidermin precursor peptide, *J. Bacteriol. 178*, 284-288.
21. Abts, A., Montalban-Lopez, M., Kuipers, O. P., Smits, S. H., and Schmitt, L. (2013) NisC binds the FxLx motif of the nisin leader peptide, *Biochemistry 52*, 5387-5395.
22. Xu, Y., Li, X., Li, R., Li, S., Ni, H., Wang, H., Xu, H., Zhou, W., Saris, P. E., Yang, W., Qiao, M., and Rao, Z. (2014) Structure of the nisin leader peptidase NisP revealing a C-terminal autocleavage activity, *Acta Crystallogr. Sect. D Biol. Crystallogr. 70*, 1499-1505.
23. Ortega, M. A., Velásquez, J. E., Garg, N., Zhang, Q., Joyce, R. E., Nair, S. K., and van der Donk, W. A. (2014) Substrate specificity of the lanthipeptide peptidase ElxP and the oxidoreductase ElxO, *ACS Chem. Biol. 9*, 1718-1725.
24. Young, C. L., Britton, Z. T., and Robinson, A. S. (2012) Recombinant protein expression and purification: a comprehensive review of affinity tags and microbial applications, *Biotechnol. J. 7*, 620-634.

25. Begley, M., Cotter, P. D., Hill, C., and Ross, R. P. (2009) Identification of a novel two-peptide lantibiotic, lichenicidin, following rational genome mining for LanM proteins, *Appl. Environ. Microbiol. 75*, 5451-5460.
26. Kuhn, M. L., Prachi, P., Minasov, G., Shuvalova, L., Ruan, J., Dubrovska, I., Winsor, J., Giraldi, M., Biagini, M., Liberatori, S., Savino, S., Bagnoli, F., Anderson, W. F., and Grandi, G. (2014) Structure and protective efficacy of the *Staphylococcus aureus* autocleaving protease EpiP, *FASEB J. 28*, 1780-1793.
27. Zhang, Q., Ortega, M., Shi, Y., Wang, H., Melby, J. O., Tang, W., Mitchell, D. A., and van der Donk, W. A. (2014) Structural investigation of ribosomally synthesized natural products by hypothetical structure enumeration and evaluation using tandem MS, *Proc. Natl Acad. Sci. U.S.A. 111*, 12031-12036.
28. Aon, J. C., Caimi, R. J., Taylor, A. H., Lu, Q., Oluboyede, F., Dally, J., Kessler, M. D., Kerrigan, J. J., Lewis, T. S., Wysocki, L. A., and Patel, P. S. (2008) Suppressing posttranslational gluconoylation of heterologous proteins by metabolic engineering of Escherichia coli, *Appl. Environ. Microbiol. 74*, 950-958.
29. Bindman, N. A., and van der Donk, W. A. (2013) A general method for fluorescent labeling of the N-termini of lanthipeptides and its application to visualize their cellular localization, *J. Am. Chem. Soc. 135*, 10362-10371.
30. Dawson, P. E., and Kent, S. B. H. (2000) Synthesis of native proteins by chemical ligation, *Annu. Rev. Biochem. 69*, 923-960.
31. Muir, T. W. (2003) Semisynthesis of proteins by expressed protein ligation, *Annu. Rev. Biochem. 72*, 249-289.
32. Kraut, J. (1977) Serine proteases: structure and mechanism of catalysis, *Annu. Rev. Biochem. 46*, 331-358.
33. Bryan, P., Alexander, P., Strausberg, S., Schwarz, F., Lan, W., Gilliland, G., and Gallagher, D. T. (1992) Energetics of folding subtilisin BPN', *Biochemistry 31*, 4937-4945.
34. Wang, J., Zhang, L., Teng, K., Sun, S., Sun, Z., and Zhong, J. (2014) Cerecidins, novel lantibiotics from *Bacillus cereus* with potent antimicrobial activity, *Appl. Environ. Microbiol. 80*, 2633-2643.
35. Lohans, C. T., Li, J. L., and Vederas, J. C. (2014) Structure and biosynthesis of carnolysin, a homologue of enterococcal cytolysin with D-amino acids, *J. Am. Chem. Soc. 136*, 13150-13153.
36. Hiller, K., Grote, A., Scheer, M., Munch, R., and Jahn, D. (2004) PrediSi: prediction of signal peptides and their cleavage positions, *Nucleic Acids Res. 32*, 375-379.
37. Kuipers, A., de Boef, E., Rink, R., Fekken, S., Kluskens, L. D., Driessen, A. J., Leenhouts, K., Kuipers, O. P., and Moll, G. N. (2004) NisT, the transporter of the lantibiotic nisin, can transport fully modified, dehydrated, and unmodified prenisin and fusions of the leader peptide with non-lantibiotic peptides, *J. Biol. Chem. 279*, 22176-22182.
38. Koponen, O., Tolonen, M., Qiao, M., Wahlstrom, G., Helin, J., and Saris, P. E. J. (2002)

NisB is required for the dehydration and NisC for the lanthionine formation in the post-translational modification of nisin, *Microbiology 148*, 3561-3568.

39. Smith, C. A., Toogood, H. S., Baker, H. M., Daniel, R. M., and Baker, E. N. (1999) Calcium-mediated thermostability in the subtilisin superfamily: the crystal structure of Bacillus Ak.1 protease at 1.8 Å resolution, *J. Mol. Biol. 294*, 1027-1040.
40. Strausberg, S. L., Alexander, P. A., Gallagher, D. T., Gilliland, G. L., Barnett, B. L., and Bryan, P. N. (1995) Directed evolution of a subtilisin with calcium-independent stability, *Bio-Technol. 13*, 669-673.
41. Strausberg, S. L., Ruan, B., Fisher, K. E., Alexander, P. A., and Bryan, P. N. (2005) Directed coevolution of stability and catalytic activity in calcium-free subtilisin, *Biochemistry 44*, 3272-3279.
42. Siezen, R. J., de Vos, W. M., Leunissen, J. A., and Dijkstra, B. W. (1991) Homology modelling and protein engineering strategy of subtilases, the family of subtilisin-like serine proteinases, *Protein Eng. 4*, 719-737.
43. Perona, J. J., and Craik, C. S. (1995) Structural basis of substrate specificity in the serine proteases, *Protein Sci. 4*, 337-360.
44. Li, B., Cooper, L. E., and van der Donk, W. A. (2009) *In vitro* studies of lantibiotic biosynthesis, *Methods Enzymol. 458*, 533-558.
45. Shi, Y., Yang, X., Garg, N., and van der Donk, W. A. (2011) Production of lantipeptides in *Escherichia coli*, *J. Am. Chem. Soc. 133*, 2338-2341.
46. Zhang, Q., Ortega, M., Shi, Y., Wang, H., Melby, J. O., Tang, W., Mitchell, D. A., and van der Donk, W. A. (2014) Structural investigation of ribosomally synthesized natural products by hypothetical structure enumeration and evaluation using tandem MS, *Proc. Natl. Acad. Sci. U. S. A. 111*, 12031-12036.

Chapter 7. Characterization of CylA, a Protease Involved in Toxin Biosynthesis

7.1 Introduction

As described in chapter 2, cytolysin is a unique lanthipeptide produced by many clinical isolates of *Enterococcus faecalis* that displays lytic activities against eukaryotic cells including human immune cells, in addition to its antimicrobial activities. Cytolysin has been linked to virulence enhancement in animal models infected by *E. faecalis* and acute patient mortality in the clinic (*1*, *2*).

CylA is an extracellular serine protease required for the biosynthesis of cytolysin that belongs to the class II LanP family introduced in sections 1.4 and 6.1. The production of the toxin initiates with the synthesis of the linear precursor peptides $CylL_L$ and $CylL_S$ by the ribosome followed by the installation of thioether rings by the lanthionine synthetase CylM. Subsequently, CylB, a LanT protein, removes the majority of the leader peptide to generate $CylL_L$' and $CylL_S$' (**Figure 7.1**) (*3*). Extracellularly, CylA further trims these peptides by removing six amino acids at the N-terminus of the core peptides to form $CylL_L$" and $CylL_S$" (cytolysin L and S), the two peptides that make up mature cytolysin (**Figure 7.1**) (*4*). The study presented in chapter 6 suggests that after cleavage of the zymogen, the N- and C-terminal fragments are stabilized as a complex by extensive hydrophobic interactions, which is different from most subtilisin-related proteases (*5*).

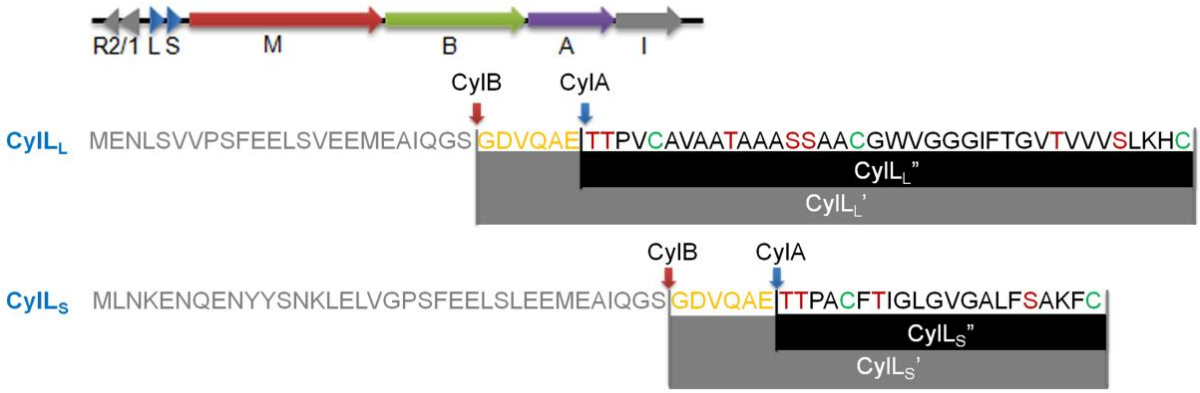

Figure 7.1 The biosynthetic gene cluster of enterococcal cytolysin and the sequential cleavage event employed during cytolysin maturation.

All currently annotated class II LanP proteins contain an N-terminal secretion signal peptide and are therefore believed to be extracellularly located (*5*). It has been confirmed for LicP that the removal of such a secretion signal peptide is achieved by an autocatalytic mechanism, with which the first 100 amino acids are tailored during maturation (*5*). Similar observations were reported for two class I LanPs, NisP and EpiP, which lack the first 195 and 99 amino acids in their mature forms, respectively (*6, 7*). The removal of a pro-sequence was suggested to activate the protease for its LanA substrate (*6*). However, to date the full length proteins have never been accessed for any of these LanPs. As a result, no activity comparison has been performed between the mature, processed form of LanP and its full length version to confirm the activation upon removal of the pro-sequence.

In order to better understand the maturation process of the enterococcal cytolysin and to mine more lanthipeptidases with novel sequence-specificities, in this chapter, I described the reconstitution of the activity of CylA *in vitro*, provide evidence that it self-activates, and demonstrate that it recognizes a specific cleavage sequence but is rather tolerant of sequence context. Similar to other sequence-specific proteases, CylA may have a broad range of applications in biotechnology (*8, 9*).

7.2 Results

The gene encoding CylA was synthesized codon-optimized for *E. coli* expression. Residues 1-26 that are predicted to constitute a secretion signal peptide were omitted. The protein was expressed in *E. coli* with an N-terminal His_6-tag and purified by immobilized metal affinity chromatography. Purified CylA showed 3 bands when analyzed by SDS-PAGE, with one band corresponding to the full length His_6-CylA-27-412 and two other bands appearing at molecular weights of about 35 kDa and 10 kDa (**Figure 7.2a**). MALDI-TOF MS analysis indicated masses of 9,591 Da and 34,815 Da (**Figure 7.2b**), consistent with an N-terminal fragment (His_6-CylA-27-95; calculated mass 9,591 Da) and a C-terminal fragment (CylA-96-412; calculated mass 34,812 Da). This observation suggested a cleavage event after Glu95, in accordance with a previous report (*4*). Moreover, His_6-CylA-27-412-E95A was expressed and purified only as the full length protein, confirming the proposed cleavage site (**Figure 7.3**).

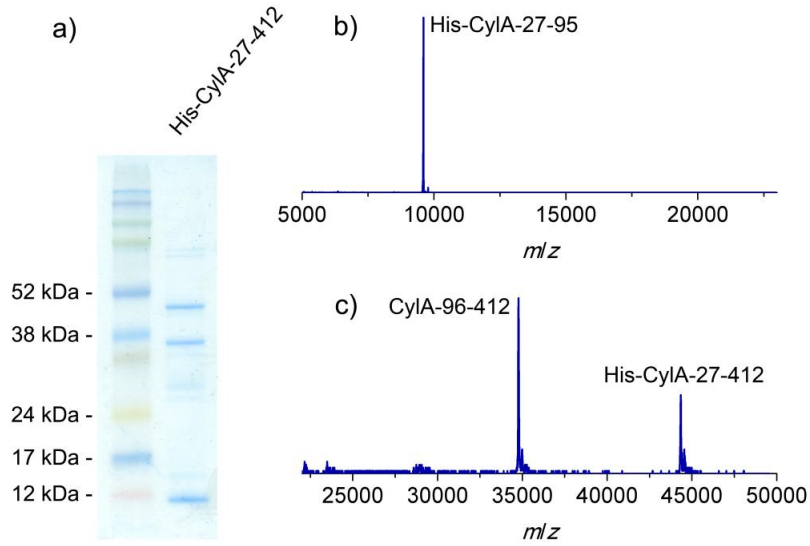

Figure 7.2 Analysis of His_6-CylA-27-412. (a) SDS-PAGE gel of His_6-CylA-27-412. MALDI-TOF mass spectra of the N-terminal fragment **(b)**, and the C-terminal fragment and full length form **(c)** of His_6-CylA-27-412.

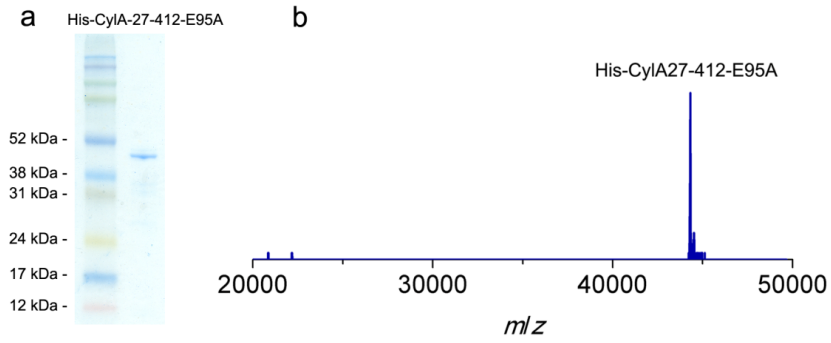

Figure 7.3 SDA-PAGE (a) and MALDI-TOF MS (b) analyses of His-CylA-27-412-E95A. (Calculated M: 44328, average mass; observed $M+H^+$: 44319, average mass.)

CylA is a subtilisin-like serine protease with a conserved catalytic triad consisting of aspartate, histidine and serine. To test whether the observed cleavage was autocatalytic, another CylA mutant was constructed with the catalytic Ser359 substituted by Ala, which was also purified only in its full length form (**Figure 7.4**). Therefore, it is concluded that CylA itself catalyzes the cleavage at position 95. These findings mirror two very recent reports, one on the class I lanthipeptidase EpiP and one on the class II peptidase LicP, both of which employ an autocatalytic mechanism to cleave a pro-sequence from the mature protease (*5, 10*). Incubation of His_6-CylA-27-412-S359A with His_6-CylA-27-412 revealed that such a cleavage could proceed intermolecularly (**Figure 7.5**).

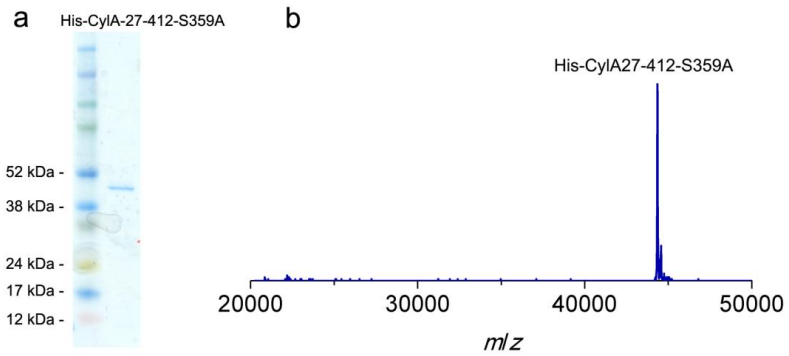

Figure 7.4 SDS-PAGE (a) and MALDI-TOF MS (b) analyses of His-CylA-27-412-S359A. (Calculated M: 44370, average mass; observed $M+H^+$: 44348, average mass.)

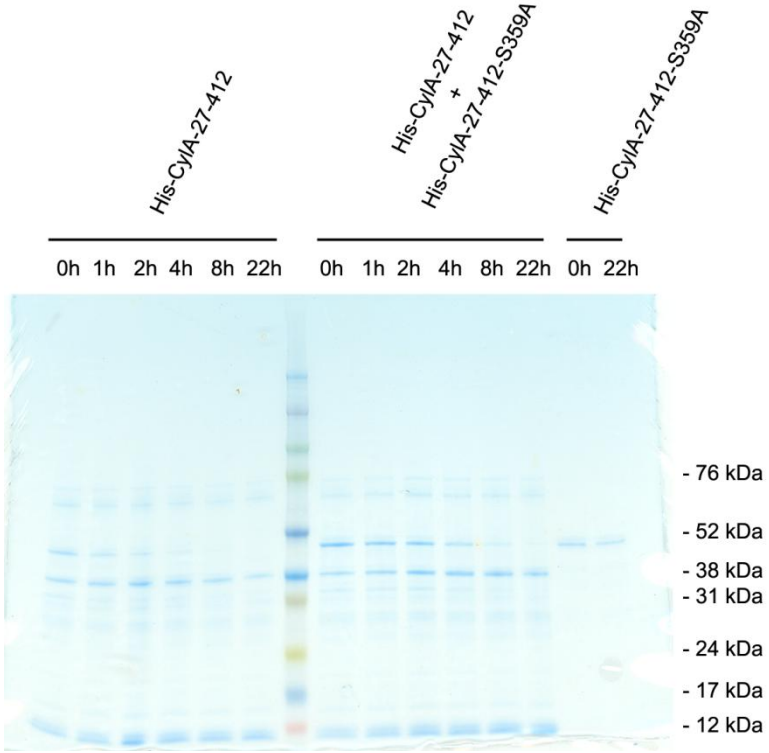

Figure 7.5 Determination if CylA cleavage is intra or intermolecular. To differentiate whether the self-cleavage of CylA occurred intramolecularly or intermolecularly, His_6-CylA-27-412 (a mixture of His_6-CylA-27-412 and partial cleaved products His_6-CylA-27-95 and CylA-95-412) and His_6-CylA-27-412-S359A were incubated alone or together. The reaction was monitored by SDS-PAGE to determine whether wild type CylA catalyzes the proteolytic cleavage of His_6-CylA-27-412-S359A. When incubated alone, His_6-CylA-27-412 self-cleaves into 2 fragments, while His_6-CylA-27-412-S359A did not show any changes throughout the 22-hour period when incubated alone. When the two proteins were incubated together, His_6-CylA-27-412-S359A was gradually consumed, suggesting that the observed cleavage of CylA can take place intermolecularly. For all reactions, His_6-CylA-27-412 was supplied with a final concentration of 0.8 mg/mL, while His_6-CylA-27-412-S359A was supplied with a final concentration of 0.2 mg/mL.

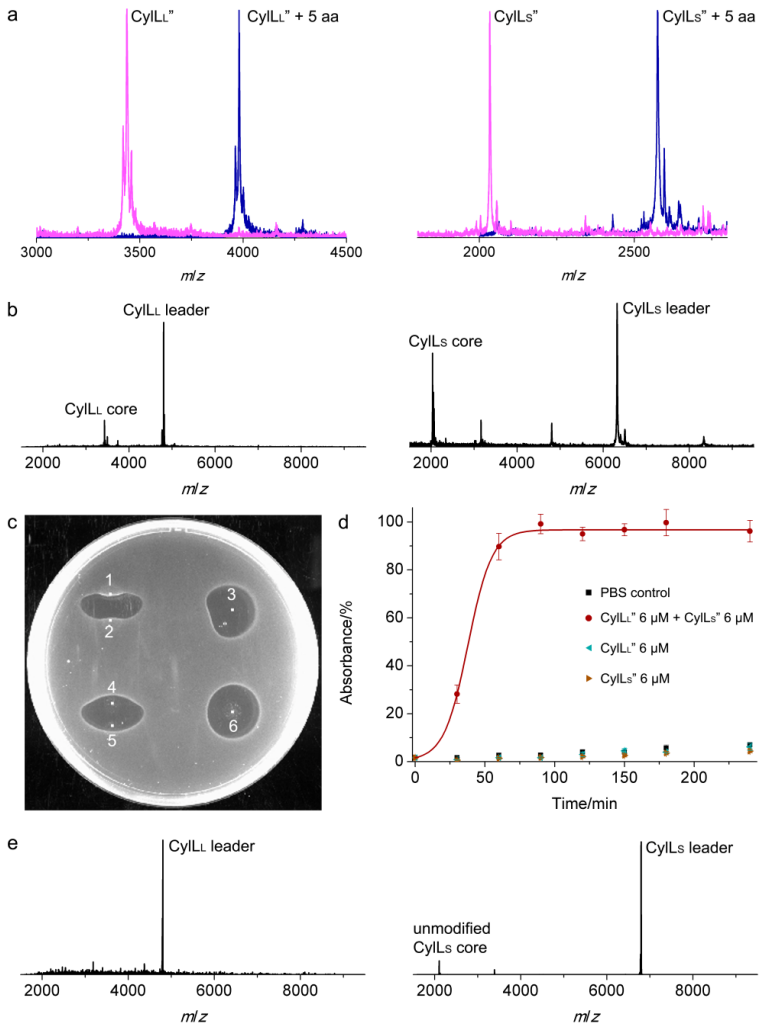

Figure 7.6 ***In vitro*** **activity of CylA. (a)** MALDI-TOF mass spectra of $CylL_L$" (left) and $CylL_S$" (right) with a 5-amino acid peptide (DVQAE) remaining from the leader incubated with (magenta) or without (blue) CylA. (DVNPE-$CylL_L$", calculated M: 3980.6, average mass; observed $M+H^+$: 3981.1, average mass. $CylL_L$", calculated M: 3438.0, average mass; observed $M+H^+$: 3438.5, average mass. DVNPE-$CylL_S$", calculated M: 2575.0, average mass; observed $M+H^+$: 2575.4, average mass. $CylL_S$", calculated M: 2032.4, average mass; observed $M+H^+$: 2034.0, average mass.) **(b)** MALDI-TOF mass spectra of modified $CylL_L$ (left) and $CylL_S$ (right) incubated with CylA. ($CylL_L$ leader, calculated M: 4806, average mass; observed $M+H^+$: 4804, average mass. $CylL_S$ leader, calculated M: 6316, average mass; observed $M+H^+$: 6323, average mass.) **(c)** Antimicrobial activities of protease digested peptides against *L. lactis* HP. Spots 1 and 2: CylM-modified and trypsin-digested $CylL_L$-E−1K and $CylL_S$-E−1K, respectively; spot 3, samples 1+2; spots 4 and 5: CylM-modified and CylA-digested $CylL_L$ and $CylL_S$, respectively; spot 6, samples 4+5. For all samples, 500 pmol were spotted. **(d)** Hemolytic activity of mature cytolysin obtained by CylA digestion. **(e)** MALDI-TOF mass spectra of linear $CylL_L$ (left) and $CylL_S$ (right) incubated with CylA. ($CylL_L$ leader, calculated M: 4806, average mass; observed $M+H^+$: 4804, average mass. $CylL_S$ leader, calculated M: 6796, average mass; observed $M+H^+$: 6796, average mass.)

I next tested the activity of CylA with the $CylL_L$ and $CylL_S$ peptides. Dehydrated and cyclized $CylL_L$ and $CylL_S$ were obtained by coexpression with their lanthionine synthetase CylM in *E. coli* (*11*). Instead of using the membrane protein CylB (*12*), the commercial protease AspN was employed, which specifically cleaved N-terminal to Asp−5, leaving five amino acids (DVQAE) on the core peptides. These peptides were incubated with CylA and the five amino acids were successfully removed (**Figure 7.6a**). I also incubated CylA with full-length modified $CylL_L$ and $CylL_S$, and MS analysis demonstrated clean removal of the entire leader peptides (**Figure 7.6b**). Cytolysin obtained in this way exhibited the anticipated antimicrobial activity against *Lactococcus lactis* HP (**Figure 7.6c**) and hemolytic activity against rabbit red blood cells (**Figure 7.6d**). Similar results were observed for lichenicidin, where the pre-treatment with LicT is not absolutely necessary for LicP recognition (*5*). Importantly, CylA did not require post-translational modifications of the precursor peptides as it also removed the leader peptides from linear $CylL_L$ and $CylL_S$ (**Figure 7.6e**).

To further understand the substrate preference of CylA with respect to linear or modified substrates, I performed a MALDI-TOF MS-based competitive assay for semi-quantitative analysis, where CylA was supplied to a mixture of equimolar amounts of modified and linear $CylL_S$ and the consumption of precursor peptides as well as the production of leader peptides was monitored over time. A Pro to Gly mutation was introduced between the hexa-histidine tag and linear $CylL_S$ peptide (G-$CylL_S$) to differentiate the otherwise identical leader peptides ($\Delta Mr = 40$ for resulting leader peptides). A 60-fold excess of modified and linear $CylL_S$ were incubated with CylA and complete consumption of modified $CylL_S$ was achieved in about 20 min. In comparison, the cleavage of linear $CylL_S$ was slower and required 160 min to complete (**Figure 7.7**). As a result, these observations suggest that, although both can be accepted as substrates, CylA prefers modified $CylL_S$ over its linear version.

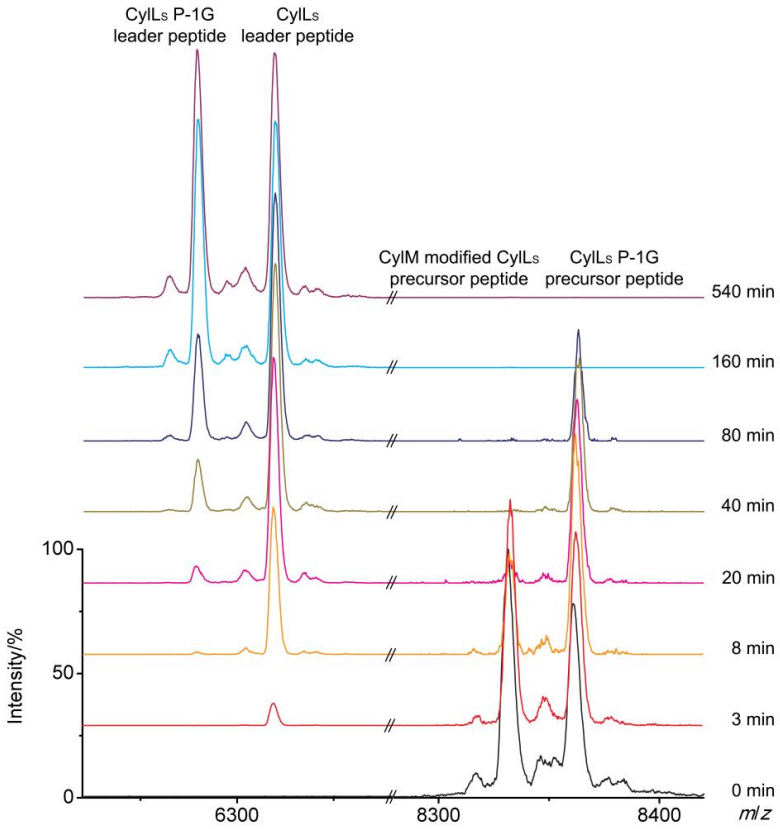

Figure 7.7 Time-dependent MALDI-TOF MS analysis of modified $CylL_S$ and linear G-$CylL_S$ treated with CylA. Peptides were each supplied with a final concentration of 12 μM in the presence of 0.2 μM His_6-CylA-27-412. The intensity of modified $CylL_S$ precursor peptide was set to 100% for starting materials in the region of 8,280-8,420 Da, whereas the intensity of $CylL_S$ P–1G leader peptide was set to 100% for reactions proceeding 160 min and 540 min (final product) in the region of 6,220-6,380 Da. For other traces, relative signal intensities in precursor peptide region and leader peptide region were adjusted to make the percentage of leader peptide peak comparing to the corresponding final product peak (percentage of formed product) and the percentage of precursor peptide peak comparing to the corresponding starting material peak (percentage of starting material left) added up to 1.

It has been suggested that sequence-specific cleavage could be generally achieved by class II lanthipeptidases with LicP as one example (*5*). To test whether this hypothesis is applicable for CylA, I engineered the proposed recognition sequence GDVQAE into two linear peptides - ProcA1.7 and NisA (**Figure 7.8**). The putative recognition sequence was installed between the leader and core peptides by substituting the residues at positions −6 to −1. Upon incubation with CylA, the leader peptides of both ProcA1.7-GDVQAE and NisA-GDVQAE were successfully removed (**Figure 7.9**).

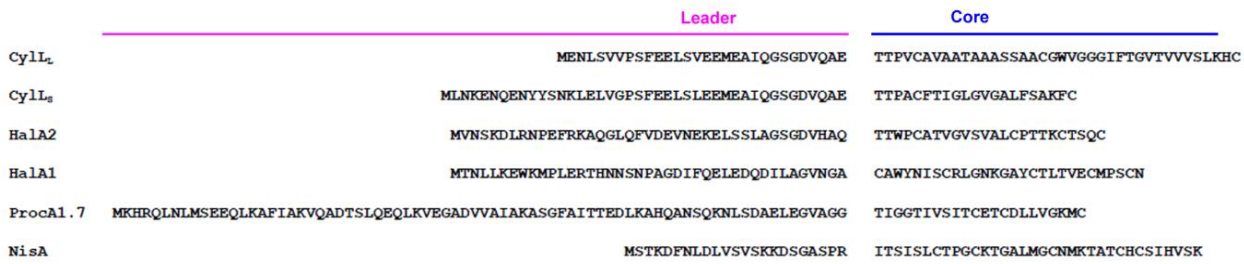

Figure 7.8 Precursor peptide sequences of lanthipeptides used in this chapter.

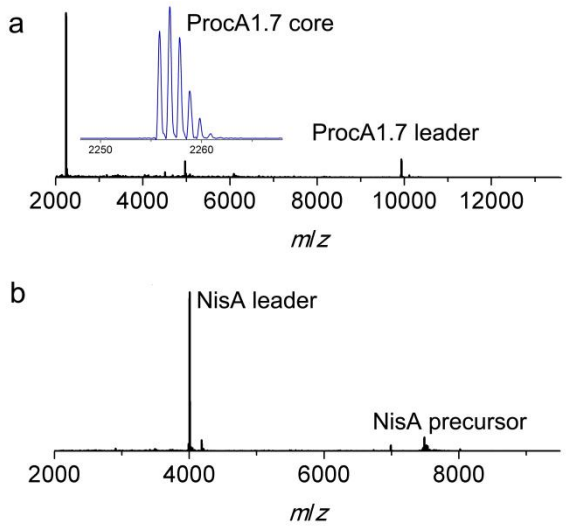

Figure 7.9 MALDI-TOF mass spectra for ProcA1.7-GDVQAE (a) and NisA-GDVQAE (b) peptides treated with CylA. (ProcA1.7 leader, calculated M: 9935, average mass; observed $M+H^+$: 9936, average mass. Oxidized ProcA1.7 core, calculated M–2H: 2254.1, isotopic mass; observed $M–2H+H^+$: 2255.8, isotopic mass. NisA leader, calculated M: 4008, average mass; observed $M+H^+$: 4005, average mass.)

The tolerance of CylA for the amino acid at the P1' position was evaluated by altering this position in ProcA1.7-GDVQAE from threonine to glycine, phenylalanine or tryptophan. CylA specifically cleaved after Glu−1 for all three mutant peptides, albeit with a lower efficiency for the T−1G mutant as demonstrated by MALDI-TOF MS (**Figure 7.10**). Peptides with N-terminal Cys residues have great utilization in native chemical ligation or expressed protein ligation (*13, 14*). However, such peptides are often not accessible by target site proteolytic cleavage as quite a few commonly used proteases do not accept a Cys at the P1' position of substrates. Yet CylA tolerates P1' Cys well by cleanly removing the leader peptides of both ProcA1.7-GDVQAE-T1C and NisA-GDVQAE-I1C peptides (**Figure 7.11**). Collectively, these results show that the activity of CylA is highly portable (**Table 7.1**).

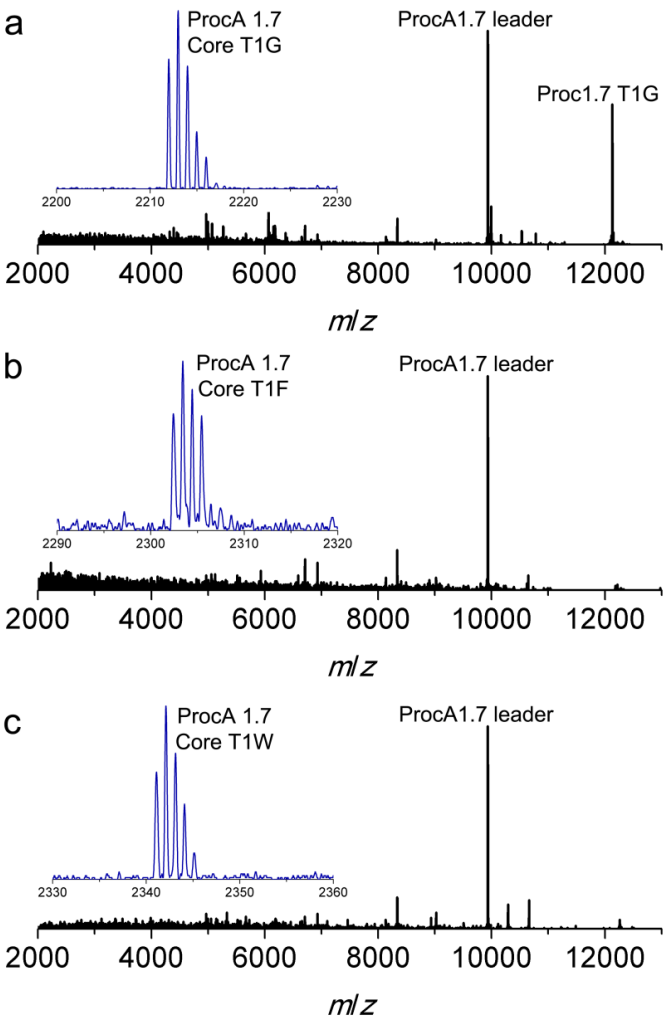

Figure 7.10 MALDI-TOF mass spectra for ProcA1.7-GDVQAE-T1G (a), ProcA1.7-GDVQAE-T1F (b), and ProcA1.7-GDVQAE-T1W (c) treated with CylA. Different instrument settings were used for the peptides shown in the insets. (ProcA1.7 leader, calculated M: 9935, average mass; observed $M+H^+$: 9936 (a), 9935 (b), 9936 (c), average masses. ProcA1.7-T1G core, calculated M: 2212.1, monoisotopic mass; observed $M+H^+$: 2212.0, monoisotopic mass. ProcA1.7-T1F core, calculated M: 2302.1, monoisotopic mass; observed $M+H^+$: 2302.4, monoisotopic mass. ProcA1.7-T1W core, calculated M: 2341.1, monoisotopic mass; observed $M+H^+$: 2341.1, monoisotopic mass.)

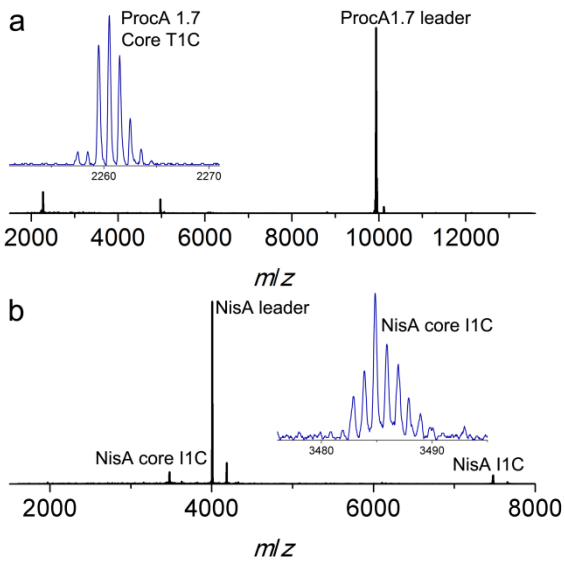

Figure 7.11 MALDI-TOF mass spectra for ProcA1.7-GDVQAE-T1C (a) and NisA-GDVQAE-T1C (b) treated with CylA. (ProcA1.7 leader, calculated M: 9935, average mass; observed $M+H^+$: 9937, average mass. ProcA1.7-T1C core, calculated M: 2258.1, monoisotopic mass; observed $M+H^+$: 2259.5, monoisotopic mass. NisA leader, calculated M: 4008, average mass; observed $M+H^+$: 4006, average mass. NisA-T1C core, calculated M: 3485.6, monoisotopic mass; observed $M+H^+$: 3482.9, monoisotopic mass.)

To expand the possible utility of CylA for mining novel RiPPs, I tested CylA's tolerance with post-translationally modified peptides. The GDVQAE sequence was engineered into HalA1 and HalA2, the precursor peptides for haloduracin α and β (Halα and Halβ), which constitute a two-component lantibiotic (*15*). HalA1-GDVQAE and HalA2-GDVQAE peptides were co-expressed with their cognate lanthionine synthetases HalM1 or HalM2 in *E. coli*, resulting in the anticipated post-translational modifications including three thioether crosslinks for HalA1 and four such rings in addition to three extra dehydrations in HalA2 (**Figures 7.12** and **7.13**). These modified peptides were then incubated with CylA and their leader peptides were successfully removed (**Figures 7.12** and **7.13**), suggesting that unnatural amino acids located in close proximity to the cleavage site do not significantly decrease CylA's proteolytic activity. Encouraged by these results, I further engineered the cleavage site between the hexa-histidine tag and modified ProcA1.7 peptide (*16*). Incubation with CylA indeed resulted in clean removal of the hexa-histidine sequence (**Figure 7.14**), confirming that CylA could be applied for expression tag removal purposes.

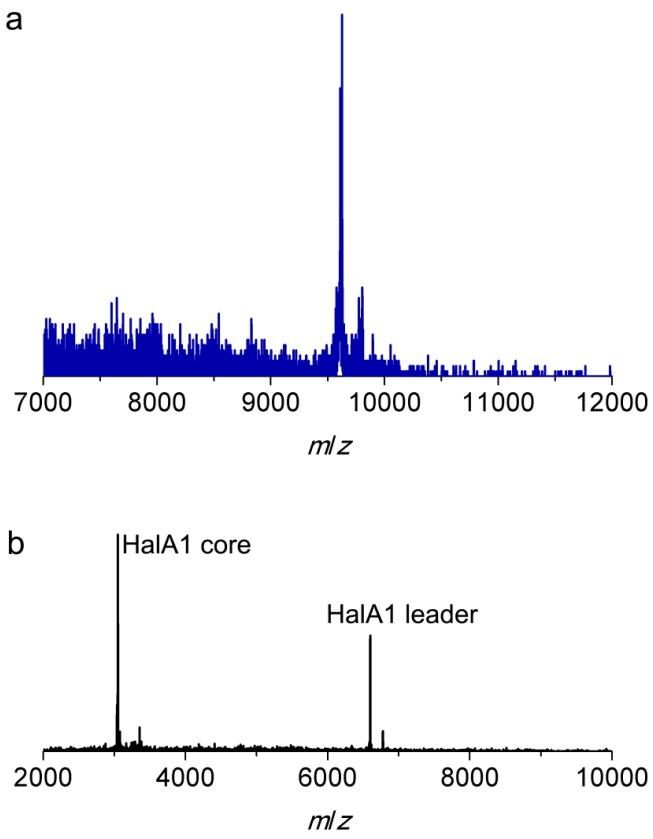

Figure 7.12 MALDI-TOF mass spectra for HalM1-modified HalA1-GDVQAE peptide with (b) or without (a) the treatment of CylA. (HalM1 modified HalA1-GDVQAE, calculated M: 9626 (with a disulfide linkage), average mass; observed $M+H^+$: 9625, average mass. HalA1 leader, calculated M: 6598, average mass; observed $M+H^+$: 6599, average mass. HalA1 core, calculated M: 3048 (reduced), average mass; observed $M+H^+$: 3048, average mass.)

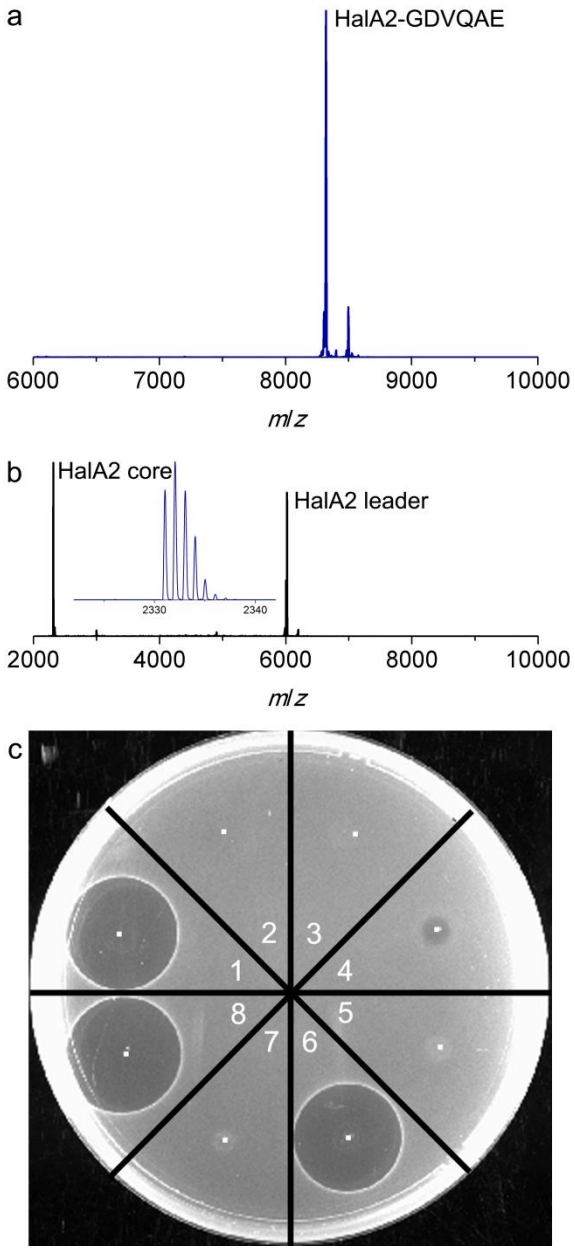

Figure 7.13 Leader peptide removal of HalA2-GDVQAE peptide catalyzed by CylA. (a) MALDI-TOF mass spectrum for HalM2-modified HalA2-GDVQAE peptide. (HalM2 modified HalA2-GDVQAE, calculated M: 8332, average mass; observed $M+H^+$: 8333, average mass.) **(b)** MALDI-TOF mass spectrum of modified HalA2-GDVQAE peptide incubated with CylA. (HalA2 leader, calculated M: 6018, average mass; observed $M+H^+$: 6819, average mass. HalA2 core, calculated M: 2330.0, monoisotopic mass; observed $M+H^+$: 2331.0, monoisotopic mass.) **(c)** Antimicrobial activity of mature Halβ obtained by CylA in combination with Halα against *Lactococcus lactis* HP. 1, 500 pmol Halα + 500 pmol Halβ; 2, 500 pmol Halα; 3, 500 pmol Halβ; 4, 500 pmol Halα + 500 pmol HalA2-GDVQAE; 5, 500 pmol HalA2-GDVQAE; 6, 500 pmol Halα + 500 pmol HalA2-GDVQAE treated with CylA; 7, 500 pmol HalA2-GDVQAE treated with CylA; 8, 100 pmol nisin.

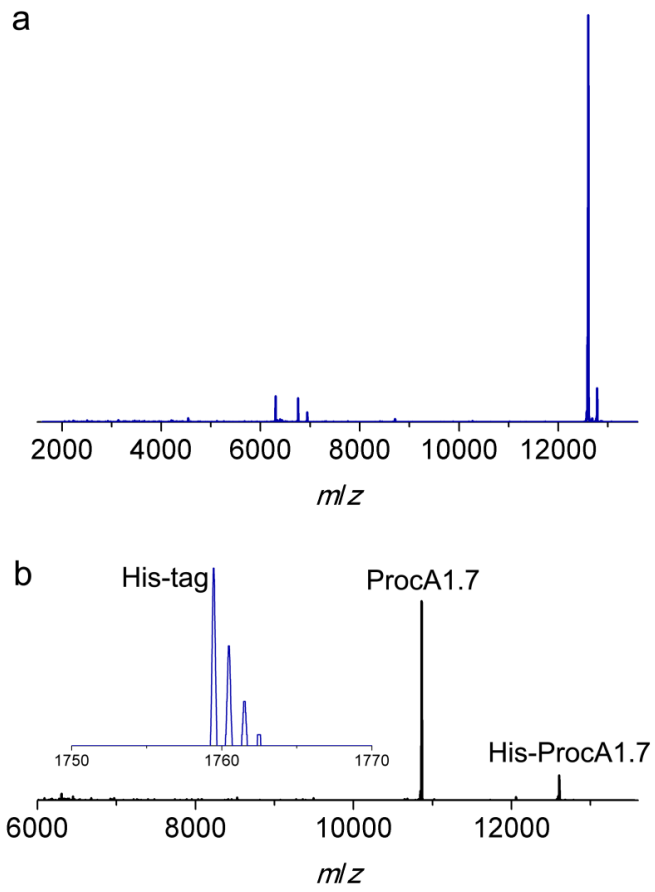

Figure 7.14 MALDI-TOF mass spectra for ProcM-modified His_6-GDVQAE-ProcA1.7 peptide with (b) or without (a) the treatment of CylA. (ProcM modified His_6-GDVQAE-ProcA1.7, calculated M: 12600, average mass; observed $M+H^+$: 12604, average mass. ProcA1.7, calculated M: 10859, average mass; observed $M+H^+$: 10862, average mass. His_6-GDVQAE, calculated M: 1758.7, monoisotopic mass; observed $M+H^+$: 1759.4, monoisotopic mass.)

I next returned to the importance of the autocatalytic processing step for activity. To assess the effect of self-cleavage on the rate of substrate cleavage, CylA-96-412 and His_6-CylA-27-412-E95A were incubated with modified $CylL_S$, and the formation of $CylL_S$'' was monitored by liquid chromatography MS. CylA-96-412 catalyzed substrate proteolysis at a rate of 20 min^{-1} under the conditions used, whereas His_6-CylA-27-412-E95A exhibited an approximate 10-fold lower rate (**Figure 7.15**). Thus, the self-cleavage event leads to activation of CylA, although the autoproteolysis is not absolutely required for protease activity under the tested conditions. The recent X-ray structure of a homolog of the class I protease EpiP illustrates that upon cleavage, the pro-domain interacts non-covalently with the catalytic domain through

complementary electrostatic surfaces (*10*). This leaves the active site of the processed enzyme more exposed, which may explain the increased activity observed in this work for processed CylA.

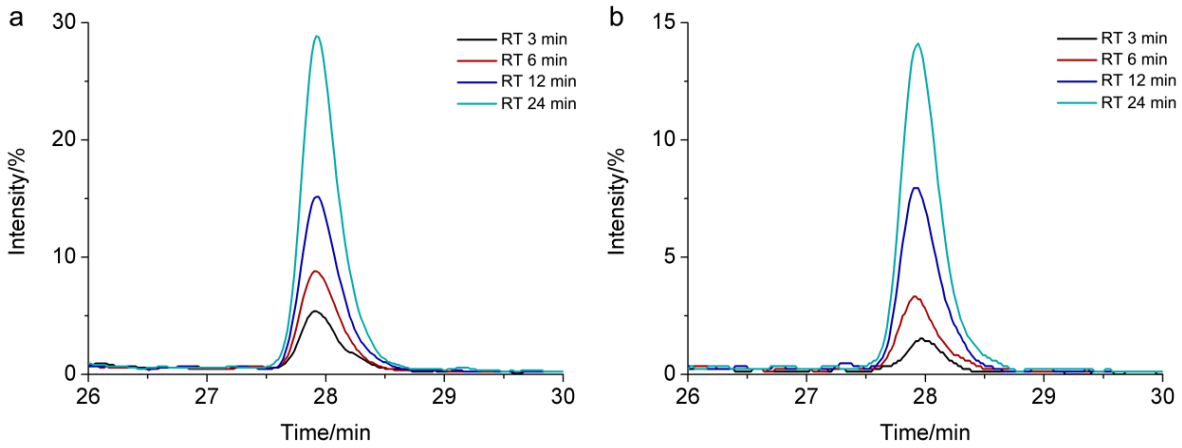

Figure 7.15 Kinetic analysis of CylA's proteolytic activities against modified CylL$_S$ with full length CylA(a), and after autoproteolytic processing (b). CylA-96-412 **(a)** and His$_6$-CylA-27-412-E95A **(b)** were at concentrations of 22 nM and 110 nM, respectively, with modified CylL$_S$ supplied at a concentration of 36 μM. Proteolytic reactions were stopped at different time points by 1% TFA and analyzed by LC/MS. Extracted ion chromatographs for mature CylL$_S$" that was produced were overlayed. Proteolytic reactions that were allowed to proceed for 24 hours were treated with the same procedure and serve as positive controls with 100% product formation.

7.3 Discussion

CylA's activity was successfully reconstituted *in vitro*, which demonstrated that it accepts both modified and linear cytolysin precursor peptides. Consistent with what has been reported for NisP, LicP and a prolyloligopeptidase-type protease FlaP identified for the biosynthesis of class III lanthipeptide flavipeptin (*5, 6, 17*), CylA cleaves modified CylL$_S$ more efficiently than linear CylL$_S$. Moreover, I provided evidence that CylA can serve as a sequence-specific protease that exhibits considerable tolerance for the P1' site and cleaves before glycine, cysteine, aromatic residues (phenylalanine, tryptophan), or branched residues (isoleucine) (**Table 7.1**), supporting the previous hypothesis regarding sequence-specific cleavage of class II LanPs (*5*). CylA can also remove small expression tags when the recognition sequence is installed. The removal of leader peptides of lanthipeptides is often difficult as most of the currently available commercial proteases do not tolerate non-proteinogenic amino acids such as dehydro amino acids,

lanthionine or methyllanthionine located in close proximity to cleavage sites. CylA does tolerate these structures and provides a significant advantage when acting on such peptides, making it a powerful tool for synthetic biology efforts with RiPPs. A general strategy for the leader peptide removal step provided by CylA will also greatly aid genome mining for novel RiPPs.

Table 7.1 Peptides with different P' sequences that are accepted by CylA.

Peptide	P' position[a]
$CylL_L$[b]	TTPVC (or MeLan)
$CylL_S$[b]	TTPAC (or MeLan)
NisA	ITSIS
ProcA1.7	TIGGT
ProcA1.7-T1G	GIGGT
ProcA1.7-T1F	FIGGT
ProcA1.7-T1W	WIGGT
NisA-I1C	CTSIS
ProcA1.7-T1C	CIGGT
HalA1	CAWYN
HalA2[c]	TTWPC (MeLan)
His_6-ProcA1.7[d]	TDPNS

[a] All peptides have the GDVQAE sequence before the P1' position.

[b] These peptides are substrates with the linear sequences shown and also after the Cys in the 5^{th} position formed a MeLan with the Thr at position 1.

[c] Peptide contains a MeLan residue at the P1' position.

[d] GDVQAE sequence is inserted between the His-tag sequence and ProcA1.7.

Although multiple groups have reported that mature LanPs purified from the supernatant of their producing strains lack the first 95-195 residues from their N-terminus (*4, 6, 7*), the results shown in this chapter serve as the first evidence in the lanthipeptide family that a LanP protease autocatalytically cleaves itself for activation (*18*). A recent study on LicP revealed a novel calcium-independent mechanism for stabilization of a subtilase, which involves the removal of a pro-sequence followed by the insertion of a tryptophan into a hydrophobic pocket in the protease domain (*5*). This model was supported by the observation that removal of the pro-sequence was essential for the production of soluble LicP. A different maturation/stabilization mechanism may be adopted by CylA, given that it can be purified as a full length protein with the pro-sequence attached. Structural information for CylA is needed to unveil the answers to these questions.

E. faecalis and *E. faecium* infections account for more than 90% of all enterococcal infections, which constitute a growing health issue worldwide (*1*). The secretion of cytolysin has been linked to the virulence of *E. faecalis*, supported by both infection model studies and clinical data (*1, 19-21*). The biosynthetic machinery that installs cytolysin could serve as a potential drug target for treating *E. faecalis* infections using an anti-virulence strategy (*1*). As the final step of cytolysin biosynthesis, the removal of GDVQAE oligopeptide from $CylL_L$' and $CylL_S$' by CylA is essential for the production of mature toxin (*4*). Drug targeting could be further assisted by the fact that CylA is secreted, unlike the other enzymes involved in cytolysin biosynthesis that require inhibitors to be membrane permeable. With heterologously expressed and purified CylA in hand, high-throughput *in vitro* screening of compound libraries can be performed to identify possible inhibitors and drug candidates.

7.4 Methods

General methods

Similar general methods and materials as what were described in chapter 5 were employed in this chapter unless specified otherwise. LC-ESI-Q/TOF MS analyses were conducted using a Micromass Q-Tof Ultima instrument (Waters) equipped with a Vydac C18 column (5 μm; 100 Å; 250 x 1.0 mm). Defibrinated rabbit blood was purchased from Hemostat Laboratories and used within 10 days of receipt. Absorbance of rabbit hemoglobin solution was measured in 96-well plates with a Synergy™ H4 Microplate Reader (BioTek). Negative numbers are used for amino acids in the leader peptide counting backwards from the leader peptide cleavage site. The gene encoding CylA was synthesized by GeneArt (Invitrogen) with codon usage optimized for *E. coli* expression. The DNA sequence of *cylA* is listed in **Table 7.2**.

Construction of pRSFDuet-1 derivatives for expression of CylA, $CylL_L$ and $CylL_S$

The *cylA*, $cylL_L$ and $cylL_S$ genes were synthesized with codon usage optimized for *E. coli* expression, amplified using appropriate primers and cloned into the MCS1 of a pRSFDuet-1 vector using restriction sites *Eco*RI and *Not*I to generate the plasmids pRSFDuet-1/CylA-27-412, pRSFDuet-1/$CylL_L$ and pRSFDuet-1/$CylL_S$. Primer sequences are listed in **Table 7.3**.

Construction of pRSFDuet-1 derivatives for expression of ProcA1.7-GDVQAE and NisA-GDVQAE peptides

Genes encoding the mutant peptides were amplified by multi-step overlap extension PCR. First, the amplification of the 5' leader part was carried out by 30 cycles of denaturing (95 °C for 10 s), annealing (55 ℃ for 30 s), and extending (72 ℃ for 15 s) using forward primers for *procA1.7* and *nisA* and appropriate reverse primers (**Table 7.3**) to generate a forward megaprimer (FMP). In parallel, PCR reactions using appropriate forward primers and reverse primers for *procA1.7* and *nisA* (**Table 7.3**) were performed to produce 3' fragments (reverse megaprimer, RMP). The 5' FMP fragment and the 3' RMP fragment were purified by 2% agarose gel followed by use of a Qiagen gel extraction kit. The 2 fragments were combined in equimolar amounts (approximately 20 ng each for a 50 μL PCR) and amplified using the same PCR conditions as above with *procA1.7* and *nisA* primers. The resulting PCR products were purified, digested and then cloned into the MCS1 of pRSFDuet-1 to generate pRSFDuet-1/ProcA1.7-GDVQAE and pRSFDuet-1/NisA-GDVQAE plasmids.

Construction of pRSFDuet-1 derivatives for expression of ProcA1.7-GDVQAE-T1G, ProcA1.7-GDVQAE-T1F, ProcA1.7-GDVQAE-T1W, ProcA1.7-GDVQAE-T1C, NisA-GDVQAE-I1C peptides and CylA-27-412-E95A, CylA-27-412-S359A proteins

Plasmids pRSFDuet-1/ProcA1.7-GDVQAE-T1G, pRSFDuet-1/ProcA1.7-GDVQAE-T1F, pRSFDuet-1/ProcA1.7-GDVQAE-T1W, pRSFDuet-1/ProcA1.7-GDVQAE-T1C, pRSFDuet-1/NisA-GDVQAE-T1C, pRSFDuet-1/CylA-27-412-E95A and pRSFDuet-1/CylA-27-412-S359A were generated using quick change methodology based on pRSFDuet-1/ProcA1.7-GDVQAE, pRSFDuet-1/NisA-GDVQAE and pRSFDuet-1/CylA-27-412, respectively. Primer sequences are listed in **Table 7.3**.

Construction of pRSFDuet-1 derivatives for co-expression of HalM1 and HalM2 with HalA1-GDVQAE and HalA2-GDVQAE, respectively

Mutant peptide genes were generated by a similar multi-step overlap extension PCR procedure as described above and cloned in the MCS1 of pRSFDuet-1/HalM1-2 and

pRSFDuet-1/HalM2-2, respectively, to generate pRSFDuet-1/HalA1-GDVQAE/HalM1-2 and pRSFDuet-1/HalA2-GDVQAE/HalM2-2 plasmids. Primer sequences are listed in **Table 7.3**.

Construction of pRSFDuet-1 derivatives for co-expression of ProcM with His_6-GDVQAE-ProcA1.7

PRSFDuet-1/His_6-GDVQAET-ProcA1.7G−1E/ProcM-2 was employed to produce modified His_6-GDVQAE-ProcA1.7 peptide and was constructed using a 2-step quick change methodology based on a previously reported plasmid pRSFDuet-1/ProcA1.7G−1E/ProcM-2(*16*, *22*). In the first quick change reaction, the gene coding GDVQAE oligopeptide was inserted between His_6-tag and ProcA sequences, resulting in pRSFDuet-1/His_6-GDVQAE-ProcA1.7G−1E/ProcM-2. The second step reaction provided the pRSFDuet-1/His_6-GDVQAET-ProcA1.7G−1E/ProcM-2 plasmid by switching the Gln residue at the P1' position to Thr. Primer sequences are listed in **Table 7.3**.

Table 7.2 The DNA sequences of *cylA* with codon usage optimized for *E. coli* expression.

cylA
ATGAAAAAACGCGGTCTGACCTATATTCTGATCAGCTATATCTTTCTGATTCTGGGCACCACCGGTTAT
GCAAGCGATCTGAGCAACAATATCAGCTTCTTTATTGATAATAGCCAGACCACCGCCATCGAAGAAAT
TGAAAGCGAACTGAGCAGCGAGAAAGTGGATTACATTCAAGAAATTGGTCTGGTGAGCTTCAAAAAC
CTGGATGATAGCGATCGCAAATTCATCGGCAAATATTTCAATGTGAGCGAAGGTAAAAAACTGCCGGA
TTTTAAACCGGAAGAAGTGAATAGCAGCATCCTGAACATTAACATCCTGAATAAAGATTTCAAAAGCT
TTAATTGGCCGTACAAAAAAATCCTGAGCCATATTGATCCGGTGAAAGAACAGCTGGGTAAAGATATT
ACCATTGCCCTGATTGATAGCGGTATTGATCGTCTGCATCCGAATCTGCAGGACAATAATCTGCGTCTG
AAAAACTATGTGAACGACATCGAACTGGATGAATATGGTCATGGCACCCAGGTTGCCGGTGTTATTGA
TACCATTGCACCGCGTGTTAATCTGAACAGCTATAAAGTTATGGATGGCACCGATGGCAATAGCATTA
ATATGCTGAAAGCAATTGTGGATGCCACCAATGATCAGGTGGATATTATCAATGTTAGCCTGGGCAGC
TACAAAAACATGGAAATTGATGATGAACGCTTTACCGTTGAAGCCTTTCGTAAAGTTGTTAATTACGC
ACGCAAAAATAACATCCTGATTGTTGCAAGCGCAGGTAATGAAAGCCGTGATATTAGCACCGGTAACG
AAAAACACATTCCGGGTGGTCTGGAAAGCGTTATTACCGTTGGTGCAACCAAAAAAAGCGGTGATATT
GCCGATTACAGCAATTATGGTAGCAACGTGAGCATTTATGGTCCGGCAGGCGGTTATGGTGATAATTA
CAAAATCACCGGTCAGATTGATGCCCGTGAAATGATGATGACCTATTATCCGACCAGCCTGGTTAGTC
CGCTGGGTAAAGCAGCAGATTTTCCGGATGGTTATACCCTGAGCTTTGGCACCAGCCTGGCAACACCG
GAAGTTAGCGCAGCACTGGCAGCAATTATGAGCAAAAATGTGGATAACAGCAAAGACAGCAATGAAG
TTCTGAACACCCTGTTTGAAAATGCCGATAGCTTCATCGATAAAAACAGCATGCTGAAATACAAAGAA
GTGCGCATTAAATAA

Table 7.3 Primer sequences for cloning of *cylA-27-412, cylA-27-412-E95A, cylA-27-412-S359A, cylL$_L$, cylL$_S$, procA1.7-GDVQAE, nisA-GDVQAE, nisA-GDVQAE-I1C, procA1.7-GDVQAE-T1G, procA1.7-GDVQAE-T1F, procA1.7-GDVQAE-T1W, procA1.7-GDVQAE-T1C, halA1-GDVQAE, halA2-GDVQAE* and *His$_6$-GDVQAET-procA1.7-G–1E*.

Primer Name	Primer Sequence (5'-3')
CylL$_L$_EcoRI_FP_Duet	AAAAA GAATTCG GAAAATC TGAGCGTTGT T
CylL$_L$_NotI_RP_Duet	AAAAA GCGGCCGC TTAGCAA TGTTTCAGGC T
CylL$_S$_EcoRI_FP	AAAAA GAATTCG CTGAATA AAGAAAATCA G
CylL$_S$_NotI_RP	AAAAA GCGGCCGC TTAGCAA AATTTTGCGC T
HalA1_SacI_FP	CGCCACTCGGAGCTCGATGACAAATCTTTT AAAAG
HalA1_SbfI_RP	ATAGTGATCCTGCAGGTTAGTTGCAAGAAGGCATG
HalA2_BamHI_FP	AAAAA GGATCC G ATGGTAAATT CAAAAGATTT
HalA2_HindIII_RP	AAAAA AAGCTT TTAGCACTGG CTTGTACACT
ProcA1.7_EcoRI_FP	GGT GCG AGG AAT TCG ATG AAG CAT AGA CAA CTA AAT CTG
ProcA1.7_NotI_RP	ATA ATA TCG CGG CCG CTC AGC ACA TTT TCC C
NisA_BamHI_FP	CTA GAT GGA TCC GAT GAG TAC AAA AGA TTT TAA CTT GG
NisA_HindIII_RP	CTA GAA GCT TTT ATT TGC TTA CGT GAA TAC TAC AAT G
CylA27-EcoRI_FP	AAAAA GAATTCG CTGAGCAACAATATCAGCTTC
CylA-NotI_RP	AAAAA GCGGCCGC TTA TTTAATGCGC ACTTCTTTGT A
CylA-E95A_QC_FP	CCGGATTTTAAACCG GCA GAAGTGAATAGCAGC
CylA-E95A_QC_RP	GCTG CTATTCACTT C TGC CGGTTT AAAATCCGG
CylA-S359A_QC_FP	GGCACC GCA CTG GCA ACACCGGAAGTTAGC GCAGCA
CylA- S359A_QC_RP	TGC CAG TGC GGTGCC AA AGCTCAGGGTATAACCATCCGG AAAATC
ProcA1.7-GDVQAE_FP	ACCATTGGGGGA ACCATTGTG
ProcA1.7-GDVQAE_RP	GGTTCC CCCAATGGT TTCTGCCTGA ACATCACC CAG CTCAGCATCA GACAGGT
NisA-GDVQAE_FP	ATTACAAGTATTTCGCTATGT
NisA-GDVQAE_RP	CGAAATACTT GTAAT TTCTGCCTGA ACATCACC ATCTTTC TTCGAAACAG ATA
ProcA1.7-GDVQAE-T1G_QC_FP	CAGGCAGAA GGT ATTGGG GGA ACCATTGTGTCGATAACCTGTG AG
ProcA1.7-GDVQAE-T1G_QC_RP	CCCAAT ACC T TCTGCCTG AAC ATC ACC CAG CTC AGC ATC A
ProcA1.7-GDVQAE-T1F_QC_FP	CAGGCAGAA TTC ATTGGG GGA ACCATTGTGTCGATAACCTGTG AG

Table 7.3 (cont.)

ProcA1.7-GDVQAE-T1F_QC_RP	CCCAAT GAA T TCTGCCTG AAC ATC ACC CAG CTC AGC ATC A
ProcA1.7-GDVQAE-T1W_QC_FP	AGGCAGAA TGG ATTGGG GGA ACCATTGTGTCGATAACCTGTG AG
ProcA1.7-GDVQAE-T1W_QC_RP	CCCAAT CCA T TCTGCCTG AAC ATC ACC CAG CTC AGC ATC A
ProcA1.7-GDVQAE-T1C_QC_FP	AGGCAGAA TGC ATTGGG GGA ACCATTGTGTCGATAACCTGTG AG
ProcA1.7-GDVQAE-T1C_QC_RP	CCCAAT GCA T TCTGCCTG AAC ATC ACC CAG CTC AGC ATC A
NisA-GDVQAE-T1C_QC_FP	AGGCAGAA TGC ACAAGTATT TCGCTATGTACACCCGGTTGTAAA
NisA-GDVQAE-T1C_QC_RP	AATACT TGT GCA TTCTGCCTG AACATC ACCATCTTTC TTCGAAACAG ATAC
His-GDVQAE-ProcA1.7_QC_FP	GGT GATGTTCAG GCAGAA CAGGATCCGAATTCG ATGAAGCAT
His-GDVQAE-ProcA1.7_QC_RP	TTCTGC CTG AAC ATC ACC GCT G TGGTGATGAT GGTGATGG
His-GDVQAE-Q1T-ProcA1.7_QC_FP	GCAGAA ACC GATCCGA AT TCGATGAAGCATAGACAACT AAATCTG ATG
His-GDVQAE-Q1T-ProcA1.7_QC_RP	TCGGATC GGT TTCTGC CTG AACATCACCG CTGTGGT
HalA1-GDVQAE_FP	TGCGCATGGTACAACATCAGC
HalA1-GDVQAE_RP	GTTGTACC ATGCGCA TTCTGCCTGA ACATCACC TAGAA TATCTTGGTC
HalA2-GDVQAE_FP	ACAACTTGGCCTTGCGCT
HalA2-GDVQAE_RP	GCAAGGCCAA GTTGT TTCTGCCTGA ACATCACC TGAACCA GCTAGAGA

Expression and purification of CylA and CylA mutants

E. coli BL21 (DE3) cells were transformed with pRSFDuet-1/CylA-27-412, pRSFDuet-1/CylA-27-412-E95A or pRSFDuet-1/CylA-27-412-S359A plasmids and plated on an LB plate containing 50 mg/L kanamycin. A single colony was picked and grown in 20 mL of LB with kanamycin at 37 ^{o}C for 12 h and the resulting culture was inoculated into 2 L of LB. Cells were cultured at 37 ^{o}C until the OD at 600 nm reached 0.5, cooled and IPTG was added to a final concentration of 0.1 mM. The cells were cultured at 18 ^{o}C for another 10 h before

harvesting. The cell pellet was resuspended on ice in LanP buffer (20 mM HEPES, 1 M NaCl, pH 7.5 at 25 ^{o}C) and lysed by homogenization. The lysed sample was centrifuged at 23,700×*g* for 30 min and the pellet was discarded. The supernatant was passed through 0.45-μm syringe filters and the protein was purified by immobilized metal affinity chromatography (IMAC) loaded with nickel as previously described (*23*). The proteins were generally eluted from the column at an imidazole concentration between 150 mM and 300 mM and the buffer was exchanged using GE PD-10 desalting columns pre-equilibrated with LanP buffer. Protein concentration was quantified by its absorbance at 280 nm. The extinction coefficient for His_6-CylA-27-412, His_6-CylA-27-412-E95A and His_6-CylA-27-412-S359A was calculated as 30,830 M^{-1} cm^{-1}. Aliquoted protein solutions were flash-frozen and kept at −80 ^{o}C until further usage.

Expression and purification of modified His_6-$CylL_L$, His_6-$CylL_S$, His_6-HalA1-GDVQAE, His_6-HalA2-GDVQAE and His_6-GDVQAE-ProcA1.7

Modified peptides were obtained using a similar procedure described previously using the corresponding co-expression plasmids (*11, 22*).

Expression and purification of linear His_6-$CylL_L$, His_6-$CylL_S$, His_6-ProcA1.7-GDVQAE, His_6-NisA-GDVQAE, His_6-ProcA1.7-GDVQAE-T1G, His_6-ProcA1.7-GDVQAE-T1F, His_6-ProcA1.7-GDVQAE-T1W, His_6-ProcA1.7-GDVQAE-T1C and His_6-NisA-GDVQAE-I1C

E. coli BL21 (DE3) cells were transformed with pRSFDuet-1/$CylL_L$, pRSFDuet-1/$CylL_S$, pRSFDuet-1/ProcA1.7-GDVQAE, pRSFDuet-1/NisA-GDVQAE, pRSFDuet-1/ProcA1.7-GDVQAE-T1G, pRSFDuet-1/ProcA1.7-GDVQAE-T1F, pRSFDuet-1/ProcA1.7-GDVQAE-T1W, pRSFDuet-1/ProcA1.7-GDVQAE-T1C or pRSFDuet-1/NisA-GDVQAE-T1C plasmids and plated on an LB plate containing 50 mg/L kanamycin. A single colony was picked and grown in 10 mL of LB with kanamycin at 37 ^{o}C for 12 h and the resulting culture was inoculated into 1 L of LB. Cells were cultured at 37 ^{o}C until the OD at 600 nm reached 0.5 and IPTG was added to a final concentration of 0.2 mM. The cells continued to be cultured at 37 ^{o}C for another 3 h before harvesting. The cell pellet was resuspended at room temperature in LanA start buffer (20 mM NaH_2PO_4, pH 7.5 at 25 ^{o}C, 500

mM NaCl, 0.5 mM imidazole, 20% glycerol) and lysed by sonication. The sample was centrifuged at 23,700×*g* for 30 min and the supernatant was discarded. The pellet was then resuspended in LanA buffer 1 (6 M guanidine hydrochloride, 20 mM NaH_2PO_4, pH 7.5 at 25 ^oC, 500 mM NaCl, 0.5 mM imidazole) and sonicated again. The insoluble portion was removed by centrifugation at 23,700×*g* for 30 min and the soluble portion was passed through 0.45-μm syringe filters. His-tagged peptides were purified by immobilized metal affinity chromatography (IMAC) loaded with nickel as previously described(*23*). The eluted fractions were desalted using reverse phase HPLC equipped with a Waters Delta-pak C4 column (15 μm; 300 Å; 25 x 100 mm) or a Strata XL polymeric reverse phase SPE column. Peptides were lyophilized and stored at −20 ^oC for future usage.

Intermolecular cleavage of His_6-CylA-27-412-S359A by His_6-CylA-27-412

Parallel reactions were set up with His_6-CylA-27-412 only (final protein concentration of 0.8 mg/mL in LanP buffer), His_6-CylA-27-412-S359A only (final protein concentration of 0.2 mg/mL in LanP buffer), and His_6-CylA-27-412 and His_6-CylA-27-412-S359A together (final protein concentrations of 0.8 mg/mL and 0.2 mg/mL, respectively). The three reactions were allowed to proceed at room temperature for 0, 1, 2, 4, 8 and 22 h before being stopped by addition of SDS loading buffer and boiling at 95 ^oC for 10 min

Proteolytic cleavage of the leader peptides

Targeted peptides were dissolved in H_2O to a final concentration of 3 mg/mL. To a 85 μL solution of peptides, 10 μL of 500 mM HEPES buffer (pH 7.5) was added followed by 5 μL of 0.5 mg/mL AspN protease (for modified $CylL_L$ and $CylL_S$ peptides), 0.1 mg/mL CylA protease (for modified and unmodified $CylL_L$ and $CylL_S$ peptides). For cleavage tests of the engineered GDVQAE peptides, CylA was added to a final concentration of 5 μg/mL (for ProcA1.7-GDVQAE-T1C and NisA-GDVQAE-T1C peptides), 25 μg/mL (for His_6-GDVQAE-ProcA1.7) or 10 μg/mL (for all other peptides), whereas the substrate peptide was added to a concentration of 0.3 mg/mL in 50 mM HEPES buffer (pH 7.5). TCEP (1 mM) was employed for modified HalA1-GDVQAE, ProcA1.7-GDVQAE-T1C and NisA-GDVQAE-T1C. The protease cleavage reaction mixtures were kept at 25 ^oC for 1 to 48 h

or left at 37 ^{o}C for less than 24 h. Osmotic pressure was adjusted with 150 mM NaCl if desired. The digested peptide mixture was directly used for antimicrobial and hemolytic assay.

Competition assay of CylA activity with modified and linear CylL$_S$

To a reaction vessel with 70 μL deionized H_2O, 5 μL each of 2 mg/mL modified CylL$_S$ and linear G-CylL$_S$ peptides were added (final peptide concentration 12 μM each) followed by 10 μL of 500 mM HEPES buffer (pH 7.5). Then, 10 μL of 0.1 mg/mL CylA was supplied (final protein concentration 0.2 μM) and the reaction was incubated at room temperature before being stopped by addition of formic acid to a final concentration of 1% at different time points for MS analysis.

Kinetic analysis of full length CylA and CylA-96-412 against modified CylL$_S$

As full length CylA is not available due to its self-cleavage, His-CylA-27-412-E95A was chosen to serve as a substituent of full length CylA as the self-cleavage was abolished whereas the conserved catalytic C-terminal region of CylA remained unchanged. To obtain the mature protease CylA-96-412, His-CylA-27-412 was aged at 4 ^{o}C for 12 hours to allow the self-cleavage to proceed until CylA-96-412 was the dominant peak monitored by MALDI-TOF MS. The aged protein mixture was directly used as a substituent of CylA-96-412. To test the proteases' activities, CylA-96-412 and His-CylA-27-412-E95A were supplied with a final concentration of 22 nM and 110 nM, respectively, with modified CylL$_S$ served at a concentration of 36 μM. The reactions were stopped at 3 minute, 6 minute, 12 minute and 24 minute by 1% TFA and the formation of mature CylL$_S$'' was monitored by liquid chromatography MS (LC/MS). A 5 μL volume of sample obtained from the cleavage reaction was applied to the column that was pre-equilibrated in aqueous solvent A. The solvents used for LC were: solvent A = 0.1% formic acid in 95% water / 5% acetonitrile and solvent B = 0.1% formic in 95% acetonitrile / 5% water. A solvent gradient of 0%-80% B over 30 min was employed and the fractionated sample was directly subjected to ESI-Q/TOF MS analysis. The production of core peptide was analyzed by extracted ion chromatography monitoring the desired product mass 1017 ($M+2H^+$).

Antimicrobial assay

L. lactis HP cells were grown in GM17 media under anaerobic conditions at 25 ℃ for 16 h. Agar plates were prepared by combining 15 mL of molten GM17 agar (cooled to 42 ℃) with 150 μL of dense cell culture. The seeded agar was poured into a sterile 100 mm round dish (VWR) to solidify. Peptide samples were directly spotted on the solidified agar. Plates were incubated at 30 ℃ for 16 h and the antimicrobial activity was determined by the size of the zone of growth inhibition. Halα and Halβ were obtained by factor Xa cleavage of modified HalA1Xa and HalA2 Xa peptides using a reported procedure (*22*). $CylL_L$" and $CylL_S$" were prepared using modified $CylL_L$-E−1K and $CylL_S$-E−1K peptides described previously (*11*).

Hemolytic assay for cytolysin

A sample of 1 mL of defibrinated rabbit blood was added into 20 mL of PBS in a 50 mL conical tube and mixed gently. The PBS-diluted blood sample was centrifuged at 800×*g* for 5 min at 4 ℃ and the supernatant containing lysed blood cells and released hemoglobin was discarded. The process was repeated 2 to 4 times until the supernatant was clear. The blood cells were then diluted with PBS to make a 5% solution, which was immediately used to test the hemolytic activity of the peptides. To an eppendorf tube, 50 μL of 5% washed red blood cell sample was added followed by the addition of the desired peptide samples or controls. PBS was used to adjust the final volume to 85 μL. All tubes were kept in a 37 ^{o}C incubator to allow the lytic reaction to proceed. At each time point, 8 or 10 μL of reaction mixture was taken out, diluted with 190 μL of fresh PBS and centrifuged at 800×*g* for 5 min. The supernatant (170 μL) was transferred to a new well and the absorbance was measured at 415 nm. The absorbance of prepared blood sample at each time point was analyzed in triplicate and the maximum absorbance was determined by adding 35 μL of 0.1% Triton in PBS to 50 μL of 5% blood sample and using the same analysis procedure.

7.5 References

1. Van Tyne, D., Martin, M. J., and Gilmore, M. S. (2013) Structure, function, and biology of the enterococcus faecalis cytolysin, *Toxins 5*, 895-911.
2. Daly, K. M., Cotter, P. D., Hill, C., and Ross, R. P. (2012) Lantibiotic production by pathogenic microorganisms, *Curr. Protein Pept. Sci. 13*, 509-523.
3. Cox, C. R., Coburn, P. S., and Gilmore, M. S. (2005) Enterococcal cytolysin: a novel two component peptide system that serves as a bacterial defense against eukaryotic and prokaryotic cells, *Curr. Protein Pept. Sci. 6*, 77-84.
4. Booth, M. C., Bogie, C. P., Sahl, H. G., Siezen, R. J., Hatter, K. L., and Gilmore, M. S. (1996) Structural analysis and proteolytic activation of enterococcus faecalis cytolysin, a novel lantibiotic, *Mol. Microbiol. 21*, 1175-1184.
5. Tang, W., Dong, S.-H., Nair, S. K., and van der Donk, W. A. (2015) Class II lanthipeptides harbor a pool of sequence-specific LanP proteases, *In preparation*.
6. van der Meer, J. R., Polman, J., Beerthuyzen, M. M., Siezen, R. J., Kuipers, O. P., and de Vos, W. M. (1993) Characterization of the *Lactococcus lactis* nisin A operon genes *nisP*, encoding a subtilisin-like serine protease involved in precursor processing, and *nisR*, encoding a regulatory protein involved in nisin biosynthesis, *J. Bacteriol. 175*, 2578-2588.
7. Geissler, S., Götz, F., and Kupke, T. (1996) Serine protease EpiP from *Staphylococcus epidermidis* catalyzes the processing of the epidermin precursor peptide, *J. Bacteriol. 178*, 284-288.
8. Young, C. L., Britton, Z. T., and Robinson, A. S. (2012) Recombinant protein expression and purification: A comprehensive review of affinity tags and microbial applications, *Biotechnol. J. 7*, 620-634.
9. Li, Q., Yi, L., Marek, P., and Iverson, B. L. (2013) Commercial proteases: Present and future, *FEBS Lett. 587*, 1155-1163.
10. Kuhn, M. L., Prachi, P., Minasov, G., Shuvalova, L., Ruan, J., Dubrovska, I., Winsor, J., Giraldi, M., Biagini, M., Liberatori, S., Savino, S., Bagnoli, F., Anderson, W. F., and Grandi, G. (2014) Structure and protective efficacy of the Staphylococcus aureus autocleaving protease EpiP, *FASEB J. 4*, 1780-1793.
11. Tang, W., and van der Donk, W. A. (2013) The sequence of the enterococcal cytolysin imparts unusual lanthionine stereochemistry, *Nat. Chem. Biol. 9*, 157-159.
12. Gilmore, M. S., Segarra, R. A., and Booth, M. C. (1990) An HlyB-type function is required for expression of the enteroccoccus faecalis hemolysin/bacteriocin, *Infect. Immun. 58*, 3914-3923.
13. Dawson, P. E., and Kent, S. B. H. (2000) Synthesis of native proteins by chemical ligation, *Annu. Rev. Biochem. 69*, 923-960.

14. Muir, T. W. (2003) Semisynthesis of proteins by expressed protein ligation, *Annu. Rev. Biochem. 72*, 249-289.
15. McClerren, A. L., Cooper, L. E., Quan, C., Thomas, P. M., Kelleher, N. L., and van der Donk, W. A. (2006) Discovery and *in vitro* biosynthesis of haloduracin, a two-component lantibiotic, *Proc. Natl. Acad. Sci. U. S. A. 103*, 17243-17248.
16. Tang, W., and van der Donk, W. A. (2012) Structural characterization of four prochlorosins: a novel class of lantipeptides produced by planktonic marine cyanobacteria, *Biochemistry 51*, 4271-4279.
17. Völler, G. H., Krawczyk, B., Ensle, P., and Süssmuth, R. D. (2013) Involvement and unusual substrate specificity of a prolyl oligopeptidase in class III lanthipeptide maturation, *J. Am. Chem. Soc. 135*, 7426-7429.
18. Power, S. D., Adams, R. M., and Wells, J. A. (1986) Secretion and autoproteolytic maturation of subtilisin, *Proc. Natl. Acad. Sci. U. S. A. 83*, 3096-3100.
19. Ike, Y., and Clewell, D. B. (1984) Genetic analysis of the pAD1 pheromone response in *Streptococcus faecalis*, using transposon Tn917 as an insertional mutagen, *J. Bacteriol. 158*, 777-783.
20. Huycke, M. M., Spiegel, C. A., and Gilmore, M. S. (1991) Bacteremia caused by hemolytic, high-level gentamicin-resistant *Enterococcus faecalis*, *Antimicrob. Agents Chemother. 35*, 1626-1634.
21. Chow, J. W., Thal, L. A., Perri, M. B., Vazquez, J. A., Donabedian, S. M., Clewell, D. B., and Zervos, M. J. (1993) Plasmid-associated hemolysin and aggregation substance production contribute to virulence in experimental enterococcal endocarditis, *Antimicrob. Agents Chemother. 37*, 2474-2477.
22. Shi, Y., Yang, X., Garg, N., and van der Donk, W. A. (2011) Production of lantipeptides in *Escherichia coli*, *J. Am. Chem. Soc. 133*, 2338-2341.
23. Li, B., Cooper, L. E., and van der Donk, W. A. (2009) In vitro studies of lantibiotic biosynthesis *Methods Enzymol. 458*, 533-558.

Chapter 8. Structural Characterization of Four Prochlorosins

8.1 Introduction

Prochlorosins (Pcns) are a group of lanthipeptides produced by *Prochlorococcus* (*1*). In *Prochlorococcus* MIT9313, 30 different peptide substrates are modified by a single promiscuous enzyme ProcM to form the characteristic thioether rings (*2*). The 30 precursor peptides have highly conserved leader peptides but highly diverse core peptides where the posttranslational modifications take place (**Figure 3.6**). This class of lanthionine-containing peptides is of interest not only because of the remarkable promiscuity of the enzyme, but also because the organisms carrying these genes are cyanobacteria accounting for as much as half of the chlorophyll in the tropical and subtropical oceans with a very different lifestyle than other known lantipeptide-producing organisms (*3*). They are single-celled organisms living in oligotrophic environments at very dilute concentrations, making it debatable whether prochlorosins serve as antimicrobial defenses like most known lanthipeptides (*4*).

Analysis of the Global Oceanic Survey showed that the genes for production of these lanthionine-containing peptides are widespread (*1*, *5*). Their transcription levels respond to nitrogen starvation, suggesting that the prochlorosins are functional (*1*). However, it remains unclear why the organisms produce this class of secondary metabolites and what their functions are. In order to address these questions as well as understand the remarkable substrate tolerance of ProcM, it is important to know the chemical structures of prochlorosins, including their ring topologies and the stereochemical configurations of their Lan and MeLan residues.

Reproduced in part with the permission from:
Tang, W., and van der Donk, W. A. (2012) Structural characterization of four prochlorosins: a novel class of lantipeptides produced by planktonic marine cyanobacteria, *Biochemistry 51*, 4271-4279.

Prior to the work described in this chapter, preliminary structural characterization of several prochlorosins had been carried out by tandem mass spectrometry (MS) (*1*). The presence of a thioether crosslink prohibits fragmentation of the peptide in the ring, and therefore the fragmentation pattern can be used for ring topology prediction (*1*, *6*). However, tandem MS encounters difficulties in elucidating the ring topology for structures that contain overlapping rings. In such cases, mutagenesis is usually employed to prevent the formation of one or several of the rings such that the topology of the remaining rings can be established using tandem MS. However, because ProcM showed very low substrate specificity under laboratory conditions and can generate a variety of ring structures, it is possible that the ring topology of the mutants is not that of the wild type prochlorosins. Indeed, a single point mutation in ProcA4.3 has been shown to change the original ring pattern (*1*). Therefore, tandem MS has its limitations for structure determination of prochlorosins and other complementary techniques are required. In this study, I used NMR spectroscopy to determine the ring topologies of four prochlorosins (Pcn1.1, Pcn1.7, Pcn3.3 and Pcn4.3) for which the precursor genes have been shown to be transcribed under nitrogen starvation (*1*).

In addition to questions about the ring patterns of prochlorosins, the stereochemical configurations of their Lan and MeLan residues had not been determined prior to this work. Lan and MeLan residues in most documented lanthipeptides/lantibiotics exhibit the (2*S*, 6*R*) and (2*S*, 3*S*, 6*R*) configurations, respectively (*7*), although the unusual (2*R*, 6*R*) and (2*R*, 3*R*, 6*R*) configurations have recently been identified in the entereococcal cytolysin and the β component of haloduracin in a specific substrate motif (chapter 2) (*8*). Because ProcM generates such a diversity of ring topologies, it was previously speculated that perhaps only some rings were formed enzymatically, thus pre-organizing the partially cyclized peptides for non-enzymatic cyclization to generate the final ring structures (*1*). Determination of the configuration of Lan and MeLan residues in prochlorosins could potentially serve to assess whether a ring is formed enzymatically or non-enzymatically.

Therefore, the stereochemical configuration of Lan and MeLan residues of seven prochlorosins (Pcn1.1, Pcn1.7, Pcn2.8, Pcn2.11, Pcn3.2, Pcn3.3 and Pcn4.3) was investigated in this study by GC-MS with a chiral stationary phase and compared with synthetic standards (*9*-

11). The production of milligram quantities of the selected prochlorosins was achieved by heterologous coexpression of the precursor peptides with ProcM in *E. coli* (*12*).

8.2 Results

8.2.1 Coexpression of ProcAs and ProcM in *E. coli*

In order to obtain milligram quantities of peptides required for NMR characterization, a heterologous coexpression system was employed using *E. coli* as host (*12-14*). Using a pRSFDuet vector, ProcM and His-tagged ProcAs were coexpressed and the modified precursor peptides were purified by immobilized metal affinity chromatography (IMAC) followed by RP-HPLC. Using this procedure, around 80 mg of modified prochlorosins with their leader peptides still attached could be obtained from 1 L of LB culture. The modified precursor peptides were proteolyzed by a commercial protease at an engineered cleavage site at the −1 position to remove the leader peptides and produce the mature core peptides. Of all the selected prochlorosin mutants, only the ProcA4.3G−1R and ProcA4.3G−1E mutants were not processed by ProcM, indicating that ProcA4.3 could not be mutated at the −1 position. As a result, wild type ProcA4.3 was used for coexpression in order to obtain modified peptide. The Glu at position −6 was used as a cleavage site for endoproteinase GluC to generate a core peptide with 5 residues at its N-terminus that remained from the leader peptide. Unlike ProcA4.3, ProcA1.7 and ProcA3.3 mutated at position −1 (G−1R and G−1K, respectively) were still processed by ProcM to generate the fully modified products, but either the engineered cleavage sites were close to a dehydrated amino acid (ProcA3.3) resulting in greatly reduced efficiency of leader peptide removal by endoproteinase LysC, or the selected endoproteinase caused the cleavage in the core peptide (ProcA1.7) which greatly reduced the yield. In order to obtain sufficient amounts of the modified core peptides, the Glu at position −6 was also used for these two peptides to remove most of the leader peptide with GluC endoproteinase. MALDI-TOF mass spectra of the resulting modified core peptides are shown in **Figure 8.1**.

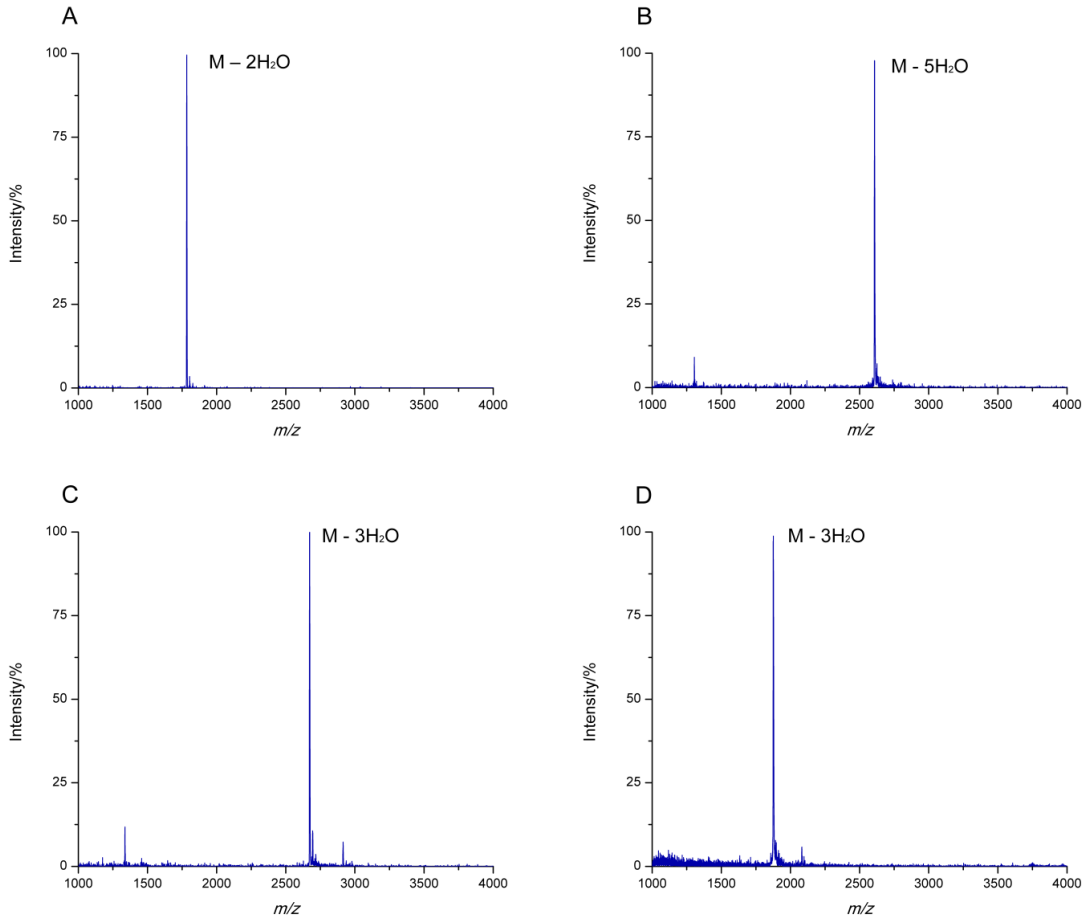

Figure 8.1 MALDI-TOF MS spectra of the core peptides of four selected prochlorosins modified by ProcM in *E. coli*. For peptide sequences, see **Figure 8.14**. (**a**) Pcn1.1. (**b**) Pcn1.7. (**c**) Pcn3.3. (**d**) Pcn4.3.

8.2.2 NMR spectral assignments

In order to determine the ring topology, the observed NMR resonances were first assigned. A series of 1D and 2D NMR experiments were performed for this purpose. The amino acid assignments were obtained primarily from TOCSY and gDQCOSY spectra. The TOCSY spectra of Pcn1.1, Pcn1.7, Pcn3.3 and Pcn4.3 are shown in **Figure 8.3**, **Figure 8.4**, **Figure 8.5** and **Figure 8.6**, respectively. The chemical shifts of amide protons in these peptides in 90% H_2O and 10% D_2O were between 7 ppm and 10 ppm, most of which showed good dispersion. The number of amide protons was determined by integration of 1D ^1H NMR spectra. Each vertical line in each TOCSY spectrum corresponds to the spin system of one amino acid, with the amide proton resonance on the x-axis and the chemical shifts of the side chain proton resonances on the y-axis.

The N-terminal residues have no amide resonance and hence were not observed. The amide signals of all amino acids were observed for each of the four prochlorosins and assigned to a specific amino acid by taking into account their characteristic chemical shifts and numbers of resonances. The amide proton signals of Dhb residues always appeared as singlets in the low field region with chemical shifts larger than 9.3 ppm, and the β/γ proton signals of Dhb displayed weak correlations to their corresponding amide protons.

Assignments of each amino acid to a specific position in the amino acid sequence were made primarily based on the connectivity of $d_{\mathrm{NN}(i,i+1)}$ as well as $d_{\alpha\mathrm{N}(i,i+1)}$ in the NOESY spectra (**Figure 8.6** for Pcn1.1, **Figure 8.7** for Pcn1.7, **Figure 8.8** for Pcn3.3, and **Figure 8.9** for Pcn4.3), taking into consideration the amino acid sequences deduced from the *procA* genes.

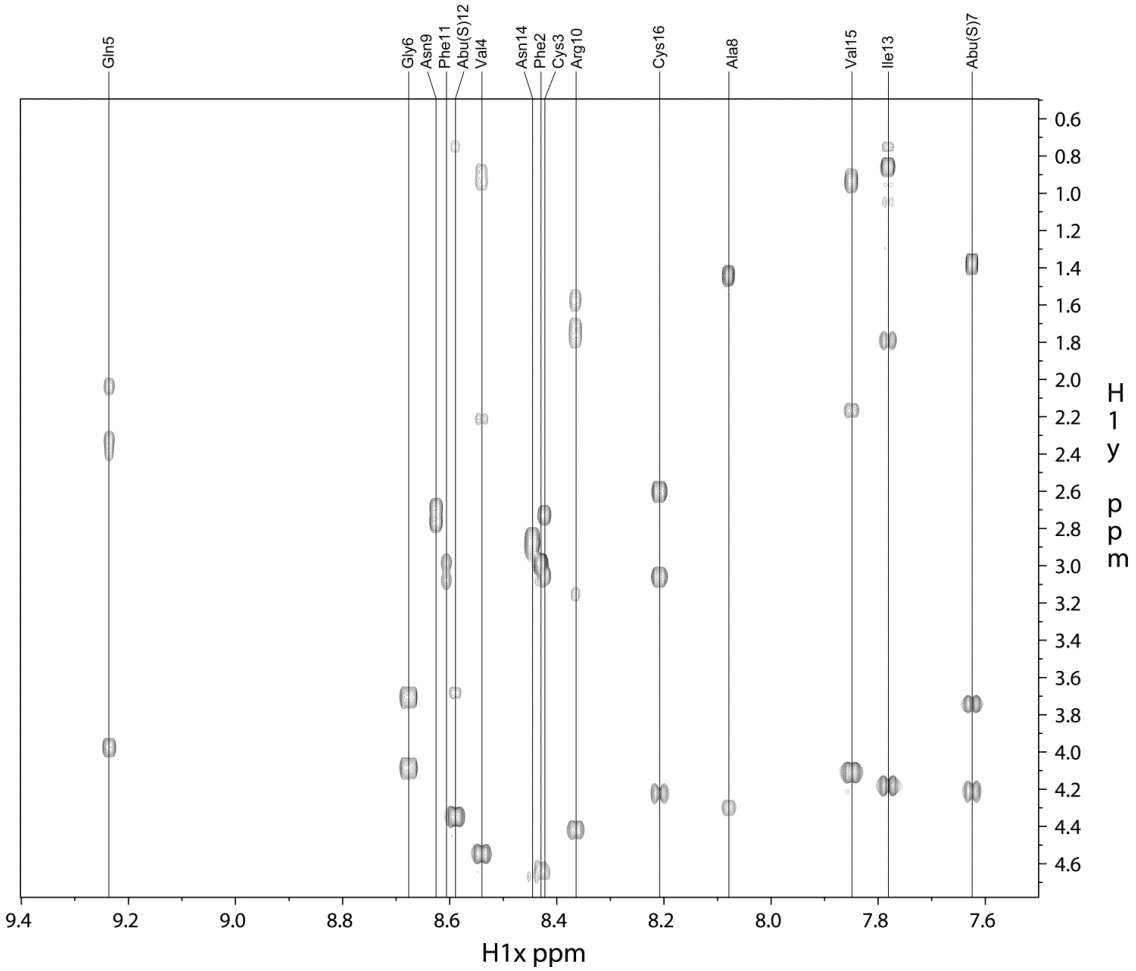

Figure 8.2 Water suppressed TOCSY spectrum identifying all residues of Pcn1.1. Each vertical line indicates a spin system corresponding to an amino acid. A water suppressed NOESY spectrum (**Figure 8.6**) along with the gene sequence was used to assign the residue numbering.

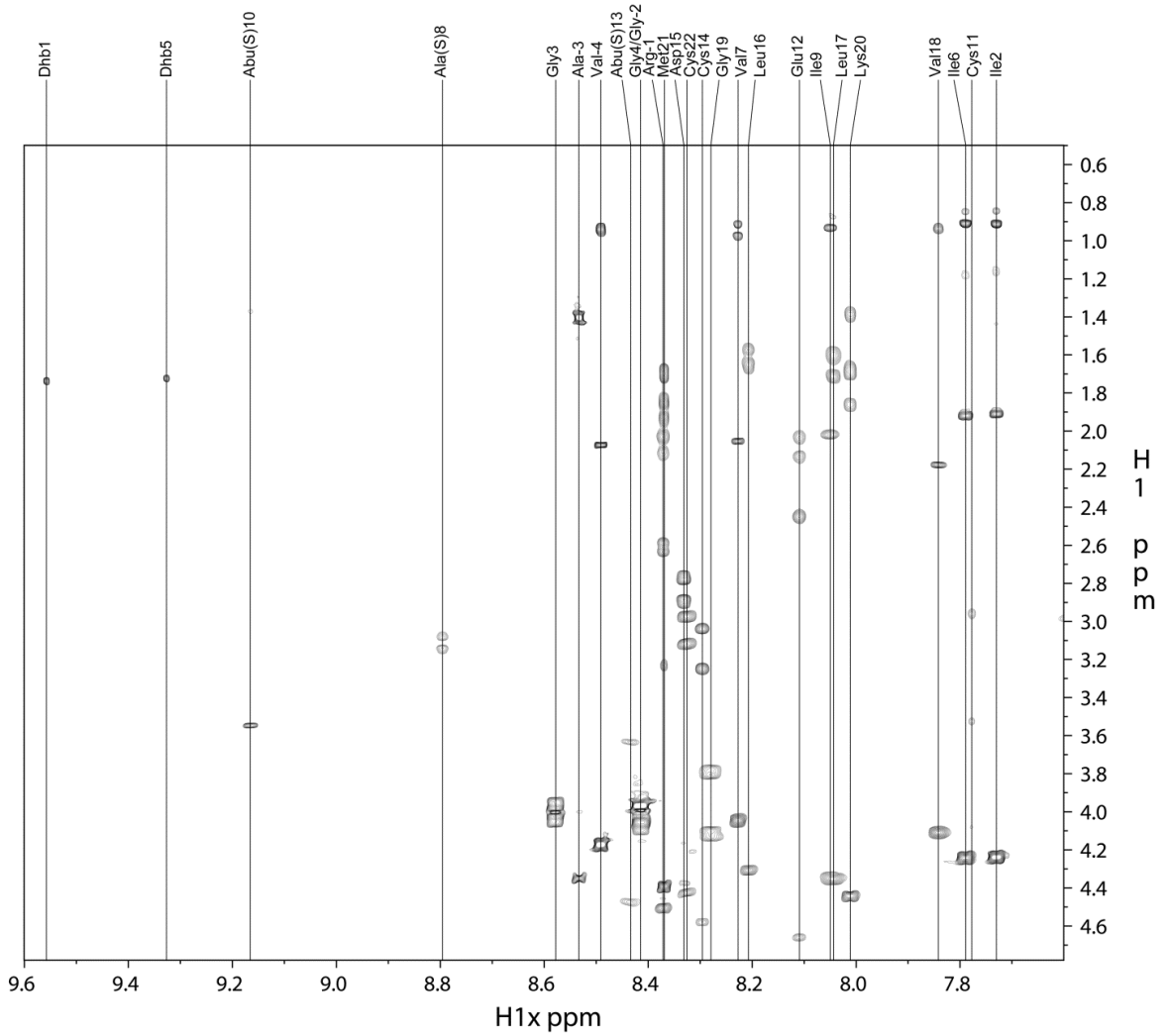

Figure 8.3 Water suppressed TOCSY spectrum identifying all residues of Pcn1.7. Each vertical line indicates a spin system corresponding to an amino acid. A water suppressed NOESY spectrum (**Figure 8.7**) along with the gene sequence was used to assign the residue numbering.

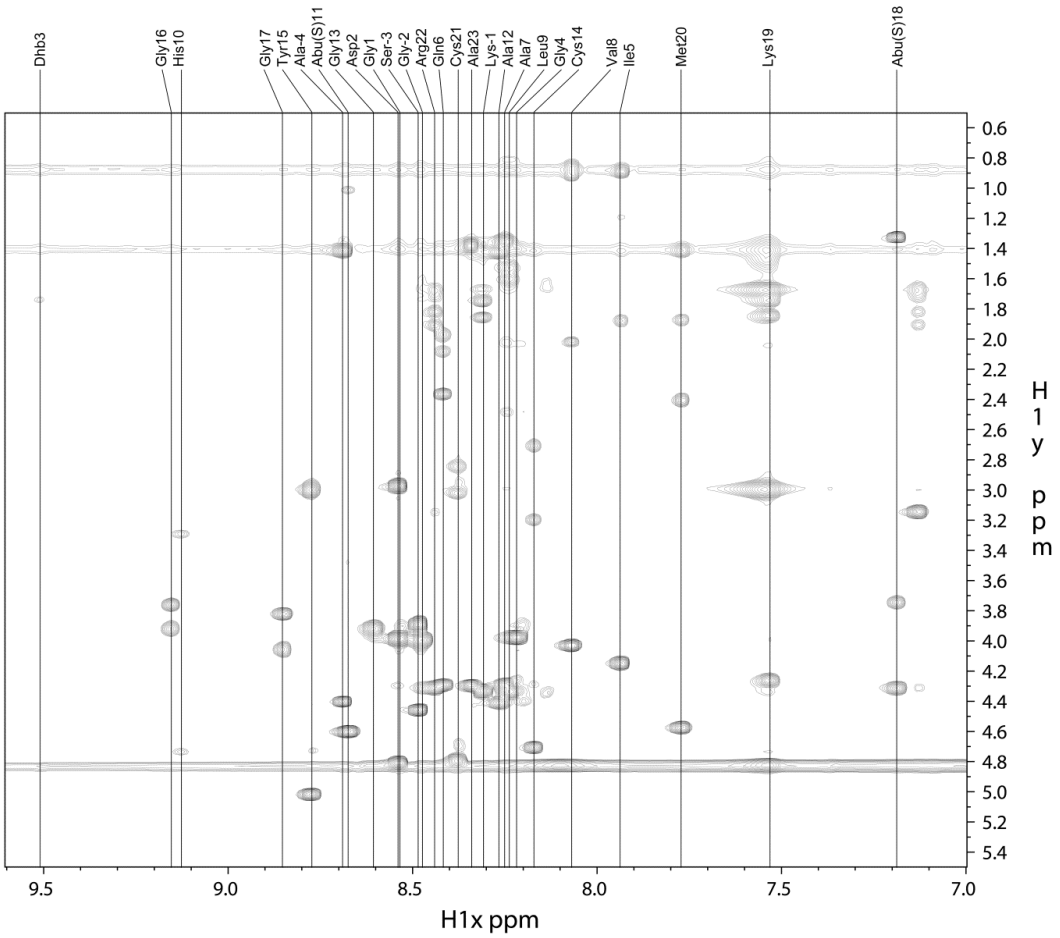

Figure 8.4 Water suppressed TOCSY spectrum identifying all amide resonances of Pcn3.3. Each vertical line indicates a spin system corresponding to an amino acid. A water suppressed NOESY spectrum (**Figure 8.8**) along with the gene sequence was used to assign the residue numbering.

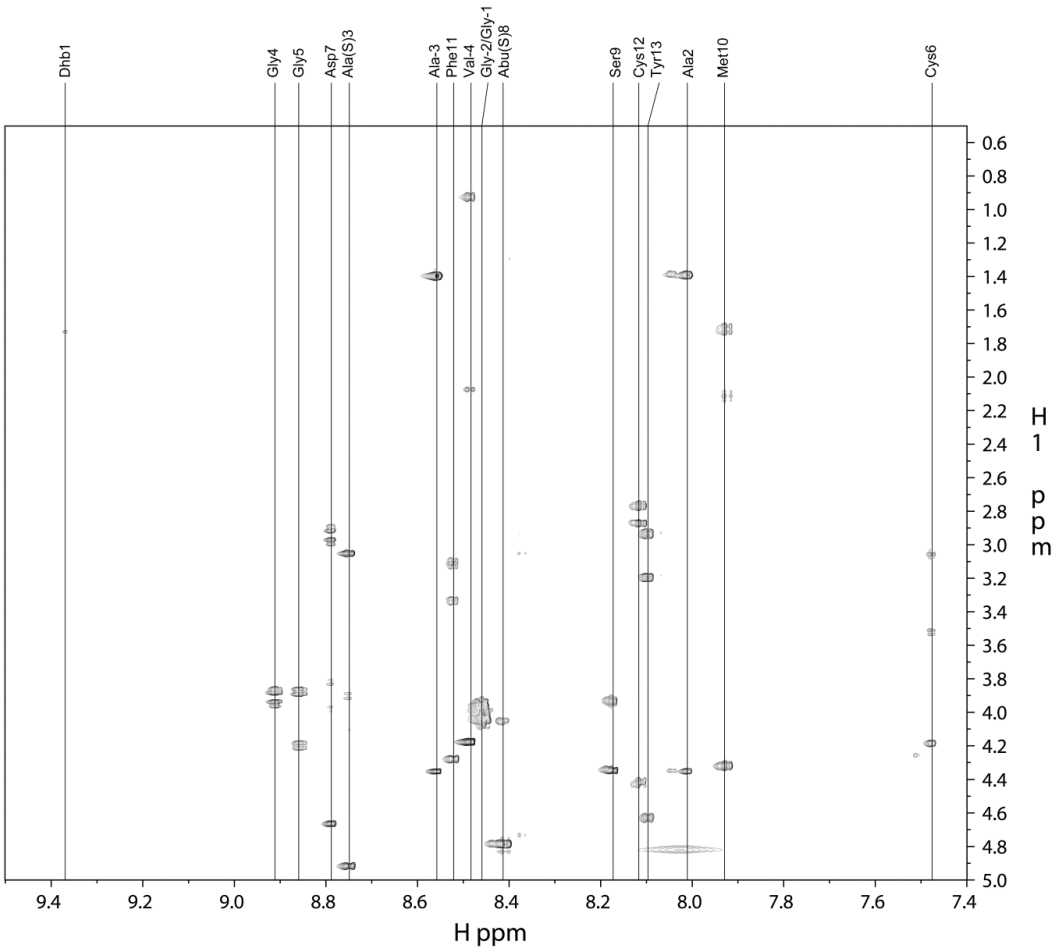

Figure 8.5 Water suppressed TOCSY spectrum identifying all amide resonances of Pcn4.3. Each vertical line indicates a spin system corresponding to an amino acid. A water suppressed NOESY spectrum (**Figure 8.9**) along with the gene sequence was used to assign the residue numbering.

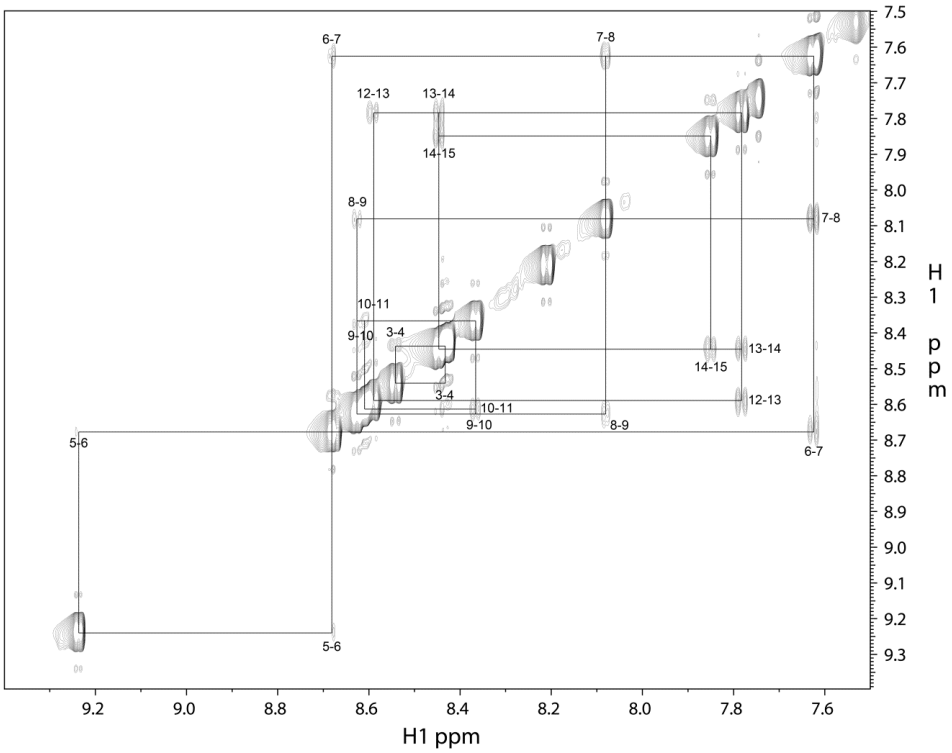

Figure 8.6 Water suppressed NOESY spectrum (mixing time = 0.20 s) for resonance amide assignments to amino acids from Pcn1.1.

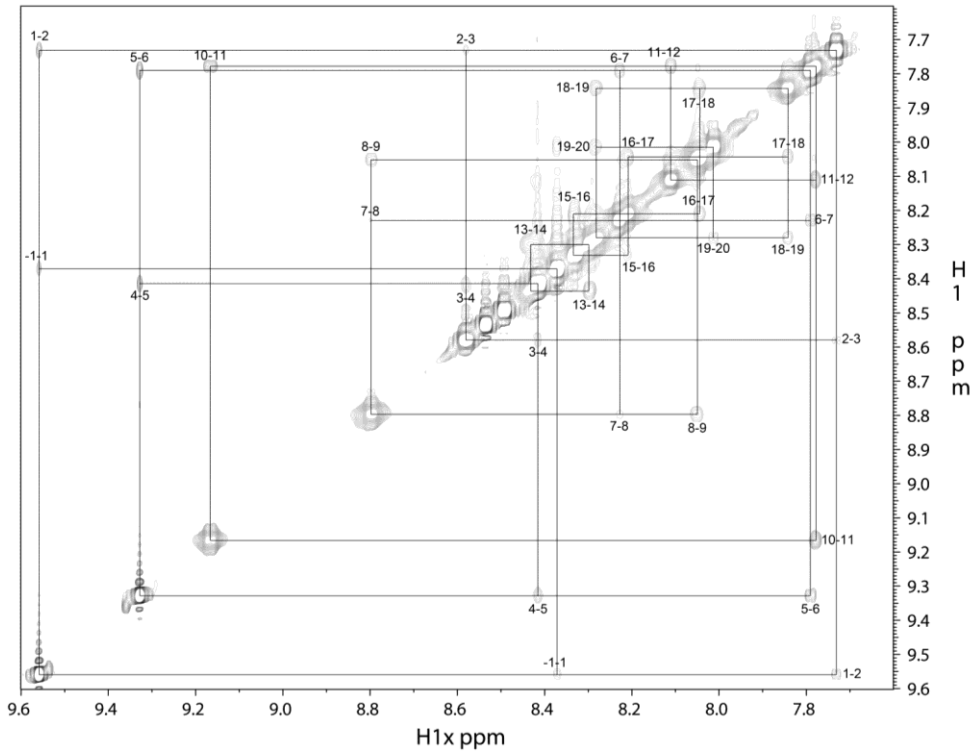

Figure 8.7 Water suppressed NOESY spectrum (mixing time = 0.20 s) for resonance amide assignments to amino acids from Pcn1.7.

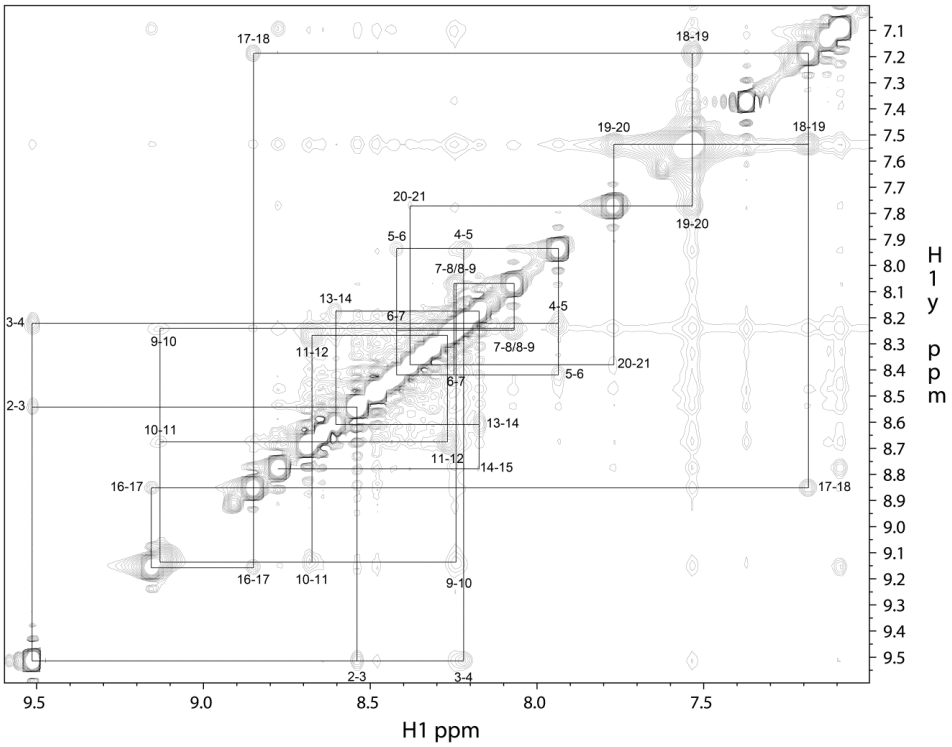

Figure 8.8 Water suppressed NOESY spectrum (mixing time = 0.20 s) for resonance amide assignments to amino acids from Pcn3.3.

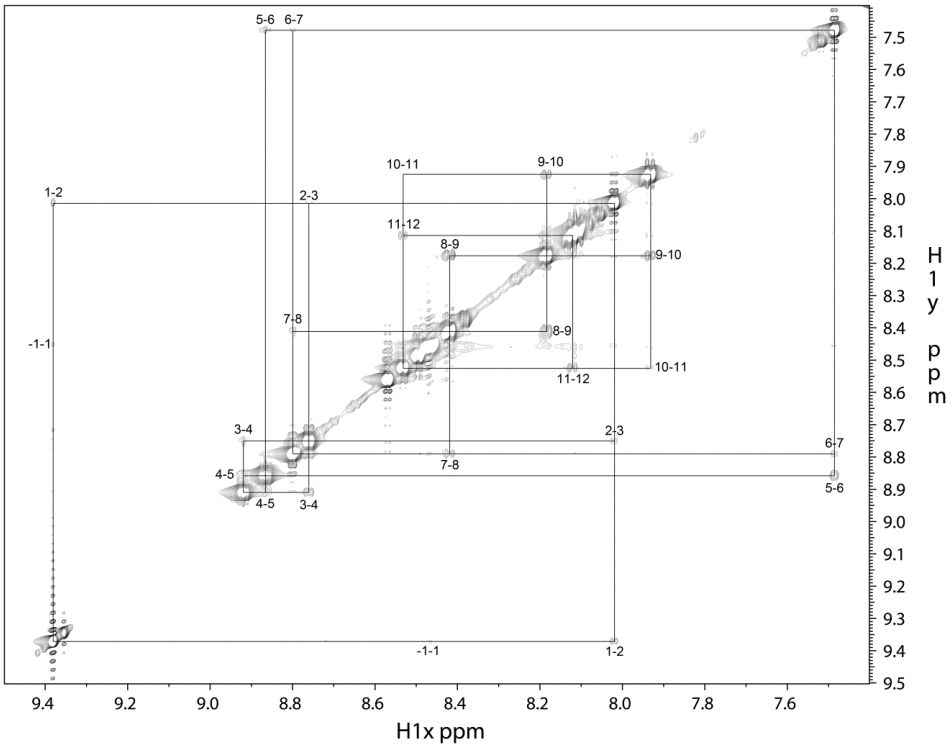

Figure 8.9 Water suppressed NOESY spectrum (mixing time = 0.20 s) for resonance amide assignments to amino acids from Pcn4.3.

8.2.3 Ring topology assigned by NMR spectroscopy

In the discussion below, the fragment of the Lan residues originating from Ser will be designated Ala(S) and the fragment that originates from Cys will be designated Cys. Similarly, the fragment of MeLan residues that originates from Thr will be designated Abu(S) (for 2-aminobutyric acid) and the fragment originating from Cys will be designated Cys. The chemical shifts for the β protons of putative Ala(S) and Abu(S) residues observed in the TOCSY experiments provided the first evidence of ring formation. Dha/Dhb residues that were converted into thioether crosslinks had chemical shifts of their β protons around 3 ppm, whereas Dha/Dhb residues that were not involved in ring formation displayed β protons chemical shifts between 6 and 7 ppm.

Next, NOESY spectra were acquired in D_2O in order to simplify the spectrum and focus on the NOEs involving the Lan/MeLan protons, which allow assignment of the ring topology (*15-20*) (**Figure 8.10** for Pcn1.7, **Figure 8.11** for Pcn1.1, **Figure 8.12** for Pcn3.3, and **Figure 8.13** for Pcn4.3). A longer NOESY mixing time of 0.40 s was used for Pcn1.7 and 4.3. Intense NOE signals between the β protons of an Ala(S)/Abu(S) and a Cys were taken as evidence for a crosslink between these residues (**Figure 8.10b**). These assignments were further supported by correlations of α protons of Ala(S)/Abu(S) to the β protons of the Cys with which it was crosslinked as well as correlations between α protons of Cys residues with the β protons of Ala(S)/Abu(S) residues (**Figure 8.10c**). In the case of MeLan, the γ-protons on the methyl groups of Abu(S) residues showed additional correlations to the α and β protons of their partner Cys (**Figure 8.10d**). For some prochlorosins (i.e. prochlorosins 1.1, 3.3 and 4.3) these latter correlations were particularly diagnostic to confirm the deduced ring structures because less signal overlap was present in this region of the spectra (**Figures 8.11-8.13**). Collectively, this approach allowed assignment of all resonances to individual residues and therefore established the topology of the rings. The ring topologies of the four selected prochlorosins were determined unambiguously (**Figure 8.14**). For Pcn1.1, Pcn1.7, and Pcn3.3, the NMR assignments agree with the previous tandem MS data (*1*). For Pcn4.3, the NMR results in this study enabled assignment of the second ring that was not possible by MS (*1*). The NMR data clearly show that this ring is formed from Thr8 and Cys12, and that Ser9 remained unmodified.

Notably, the additional amino acids that remained from the leader peptides in prochlorosins 1.7, 3.3, and 4.3 showed fewer correlation signals in the NOESY spectrum, suggesting that the N-terminus is conformationally more flexible than the modified core peptide.

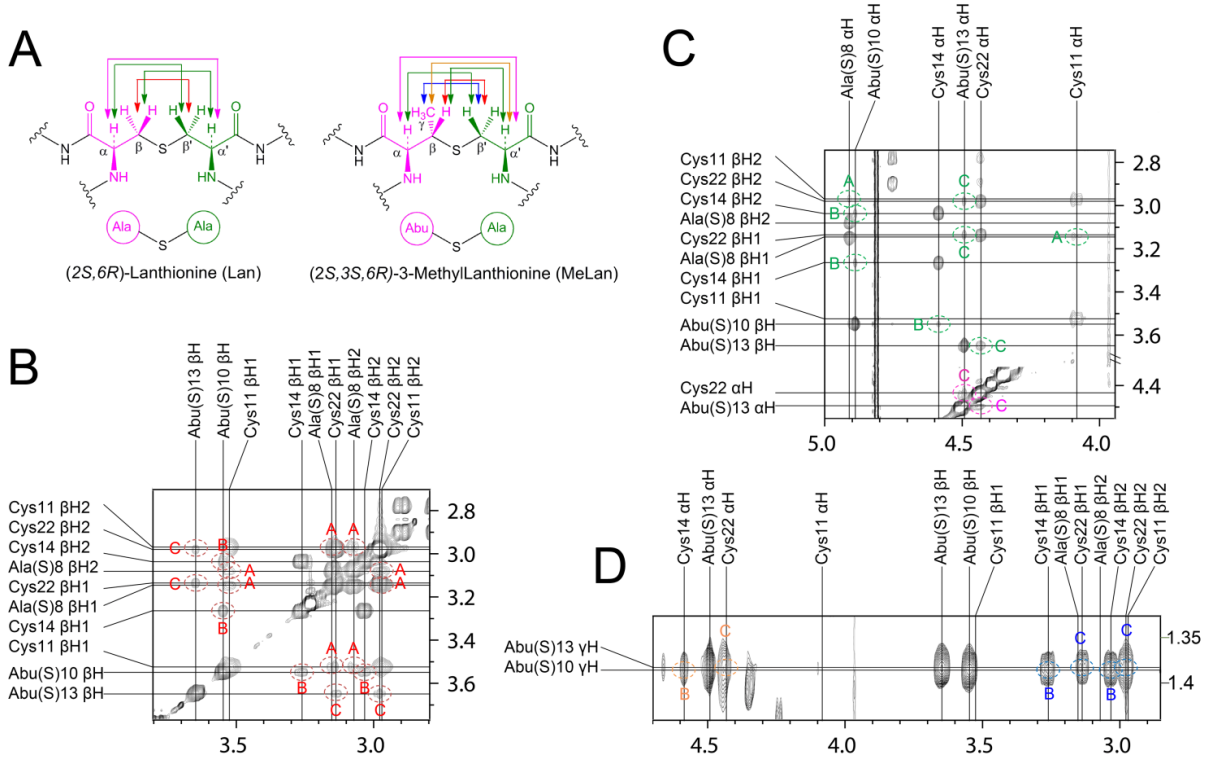

Figure 8.10 Illustration of the method used to establish thioether connectivities. (**a**) Schematic illustration of the proton correlations used to determine ring topologies. (**b**) Section of the NOESY spectrum (mixing time = 0.40 s) of Pcn1.7 in D_2O showing the correlations between β-protons. (**c**) Section of the NOESY spectrum showing the correlations of the α protons with α and β protons across the thioether bridge. (**d**) Section of the NOESY spectrum showing the correlations of the γ protons with α and β protons across the thioether bridge. In panels **b**-**d**, correlations are assigned to the A, B, and C rings of Pcn1.7 (see also **Figure 8.14**).

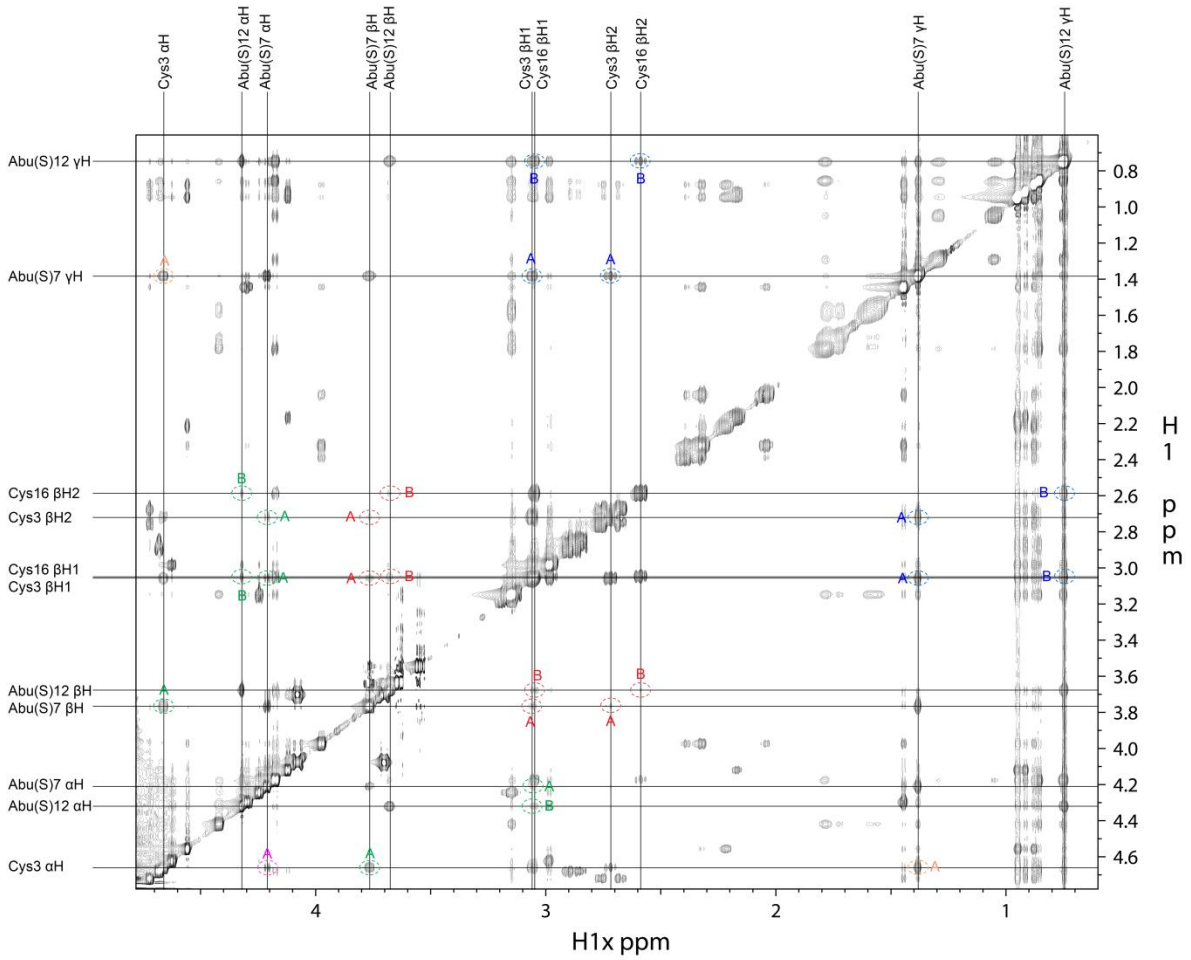

Figure 8.11 NOESY spectrum of Pcn1.1 in D_2O (mixing time = 0.20 s) for ring topology assignment. The assignments are indicated on the spectrum. Note that there are also un-assigned NOE signals that may indicate a relatively stable 3D conformation.

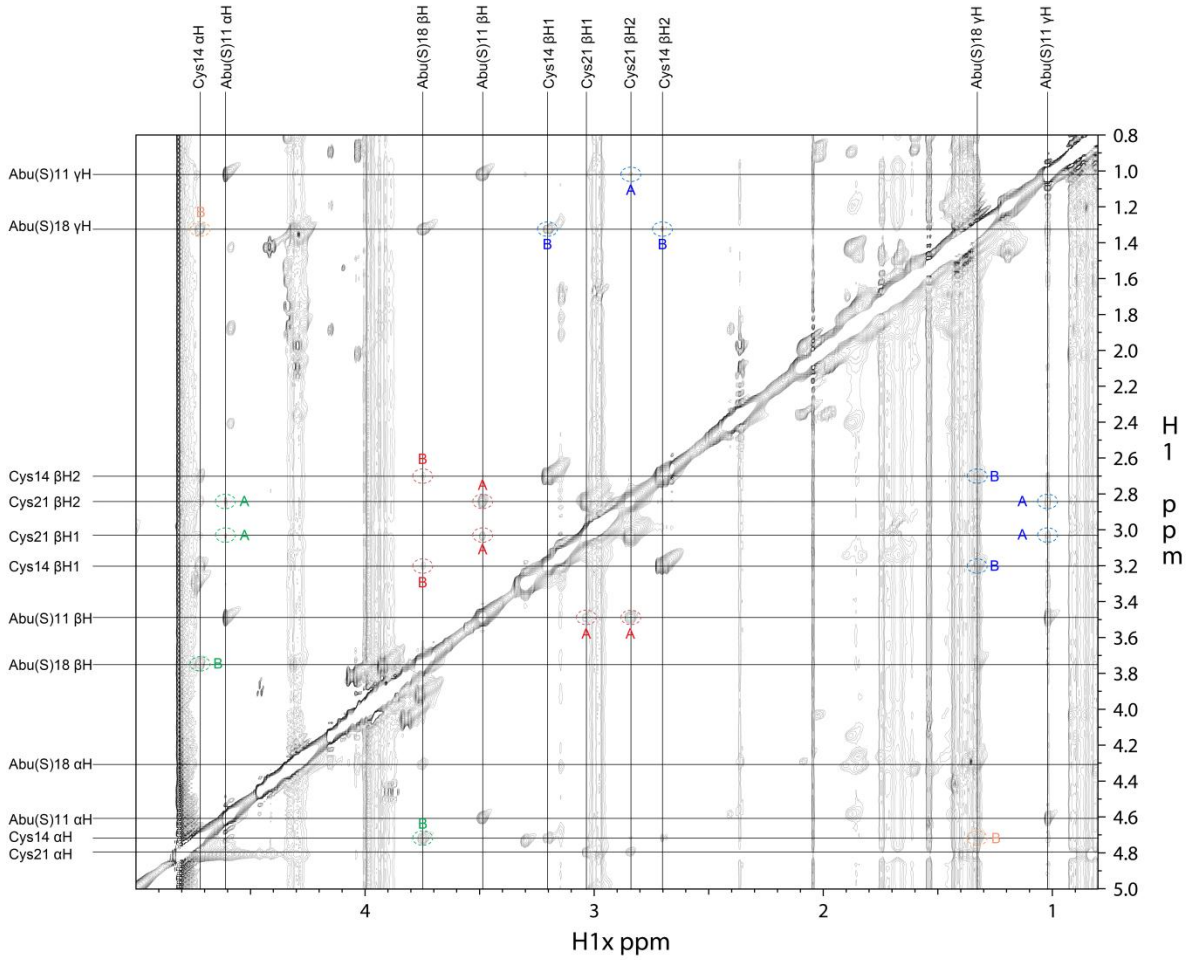

Figure 8.12 NOESY spectrum of Pcn3.3 in D_2O (mixing time = 0.20 s) for ring topology assignment. The assignments are indicated on the spectrum.

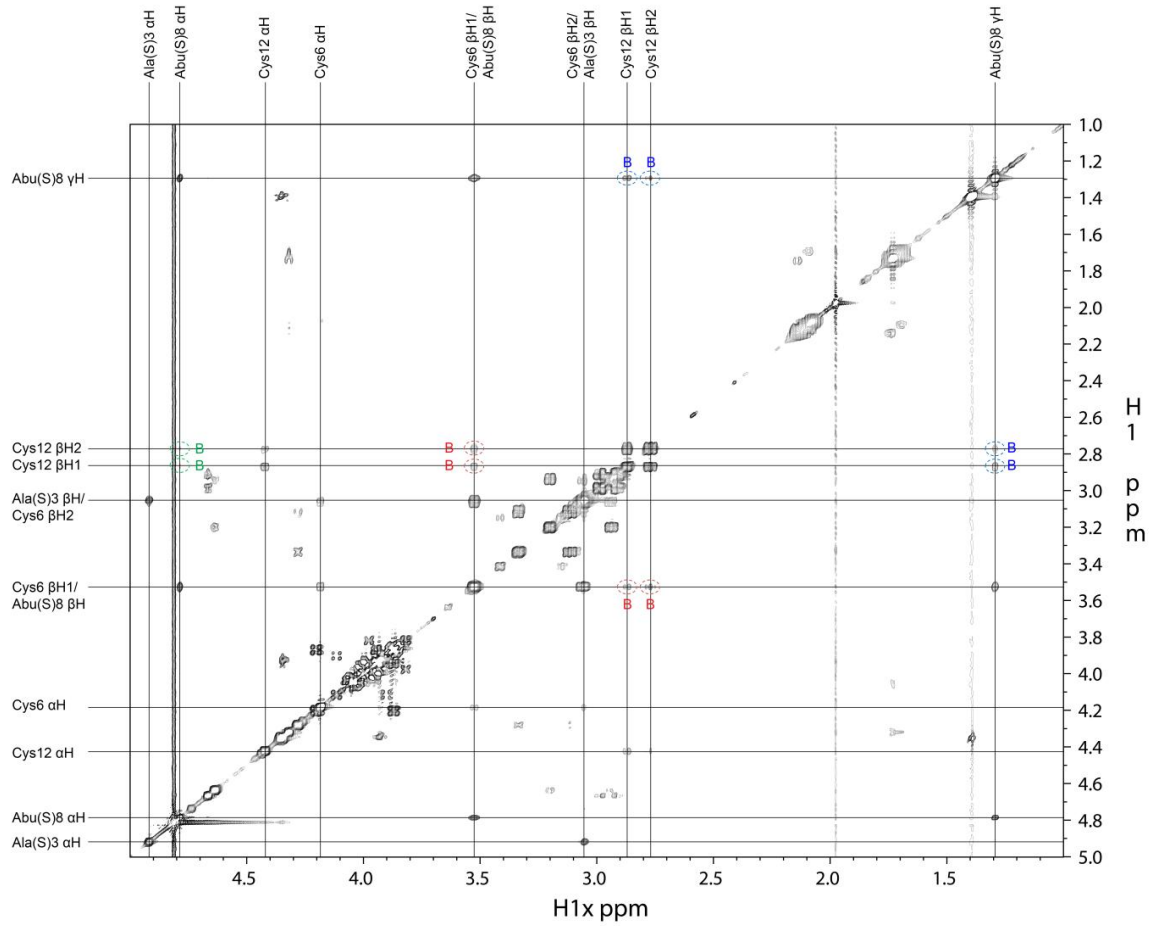

Figure 8.13 NOESY spectrum of Pcn4.3 in D2O (mixing time = 0.40 s) for ring topology assignment. The assignments are indicated on the spectrum.

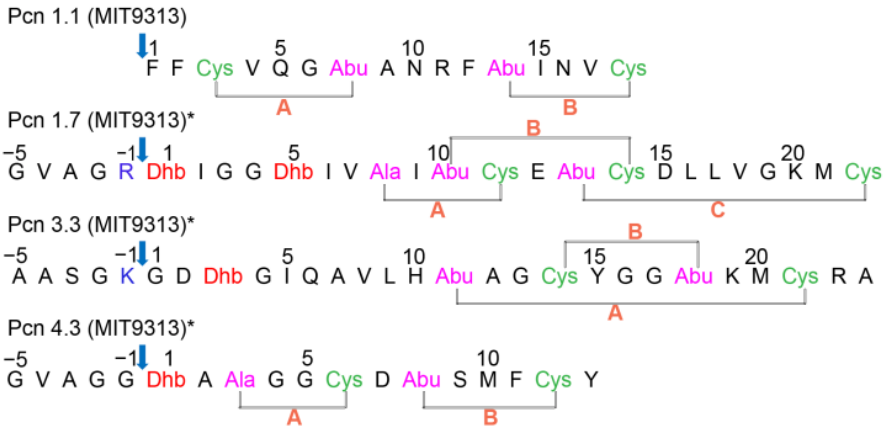

Figure 8.14 Ring topology of four selected prochlorosins assigned by NMR spectroscopy in this work. Arrows indicate the start of the putative core peptide. Asterisks indicate prochlorosins containing five more residues at their N-terminus that originate from the leader peptides. Blue residues are the engineered cleavage sites at the −1 position in the leader peptides.

8.2.4 Determination of the stereochemistry of Lan and MeLan residues

To determine the stereochemical configurations of the Lan and MeLan residues in the four prochlorosins for which the NMR structures were determined, the peptides were hydrolyzed in 6 M HCl, derivatized to their corresponding pentafluoropropionamide methyl esters, and analyzed by GC-MS with a chiral stationary phase. Lan and MeLan standards of different stereochemistry were synthesized and derivatized using previously reported methods (*10*). A Chirasil-L-Val coated GC column was used for analysis and the GC trace was monitored for a characteristic fragment of 365 Da for Lan and 379 Da for MeLan residues using electron-impact mass spectrometry. **Figure 2.5** shows the GC-MS traces of three derivatized Lan standards and two derivatized MeLan standards. The derivatized MeLan standards with DL (2*S*, 3*S*, 6*R*) and LL (2*R*, 3*S*, 6*R*) configuration were well separated on the column, with retention times of 14.1-14.4 min and 14.4-14.7 min, respectively. The derivatized Lan standards with DD (2*S*, 6*S*), DL (2*S*, 6*R*), and LL (2*R*, 6*R*) configuration eluted later than the MeLan standards, and were also well separated with retention times of 18.2-18.6 min, 18.6-19.0 min and 19.0-19.4 min, respectively. Derivatized MeLan originating from Pcn1.7 eluted with a retention time of 14.1-14.4 min (**Figure 8.15**), indicating that the MeLan in Pcn1.7 has a DL configuration. In order to further confirm this result, the derivatized DL and LL MeLan standards were co-injected with the derivatized MeLan originating from the peptide. The derivatized MeLan fragment from Pcn1.7 co-eluted with the DL MeLan standard but not with the LL MeLan standard (**Figure 8.15**). MeLan standards with the D-*allo*-L (2*S*, 3*R*, 6*R*) and L-*allo*-L (2*R*, 3*R*, 6*R*) configuration were not synthesized because the reported elution times of the derivatized MeLan diastereomers on a Chirasil-L-Val coated GC column are DL<LL<L-*allo*-L<D-*allo*-L (*10*). In this work, no derivatized MeLan residues were observed to elute after the LL standard for Pcn1.7, ruling out the presence of MeLan residues with L-*allo*-L or D-*allo*-L configuration. The MeLan residues present in the other three prochlorosins were analyzed using the same procedure and all had the DL configuration (**Figure 8.16** for Pcn1.1, **Figure 8.17** for Pcn3.3, and **Figure 8.18** for Pcn4.3). Similarly, derivatized Lan originating from Pcn1.7 co-eluted with the Lan standard with the DL configuration but not with either the DD or the LL Lan standards (**Figure 8.15**). The small shoulders on the derivatized DL-Lan peak arising from Pcn1.7 are believed to be caused by partial epimerization during HCl hydrolysis, which has been reported previously (*21*).

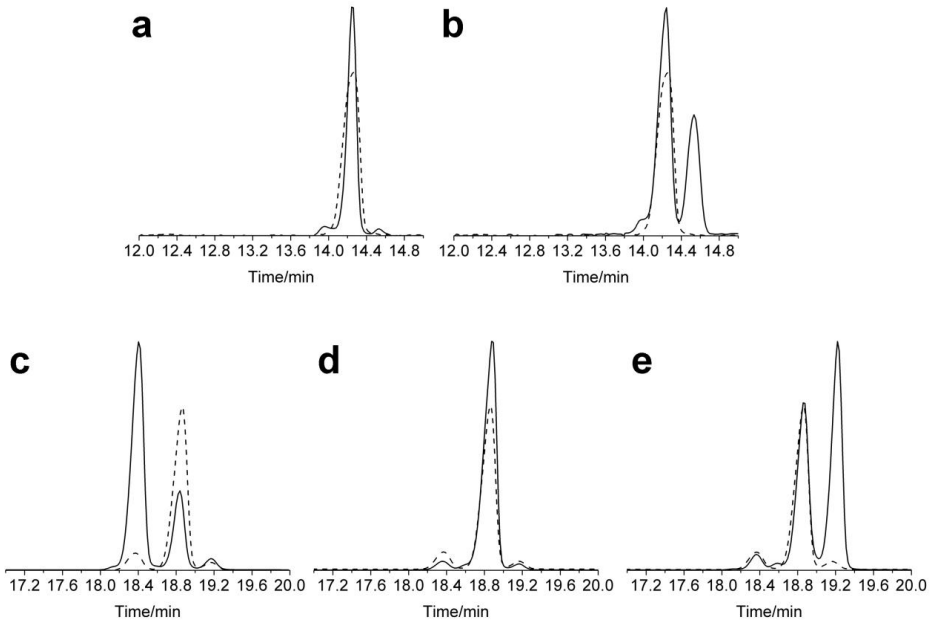

Figure 8.15 GC-MS traces (selected ion monitoring, SIM, at 365 Da for Lan and 379 Da for MeLan) of derivatized Lan/MeLan obtained from Pcn1.7 and coinjections with derivatized (Me)Lan standards. (**a**) Hydrolyzed and derivatized amino acids from Pcn1.7 (dashed line) and coinjected with DL-MeLan standard (solid line). The trace was monitored at m/z 379. D. (**b**) Hydrolyzed and derivatized amino acids from Pcn1.7 (dashed line) and coinjected with derivatized LL-MeLan standard (solid line).The trace was monitored at m/z 379. (**c**) Hydrolyzed and derivatized amino acids from Pcn1.7 (dashed line) and coinjected with DD-Lan standard (solid line). The trace was monitored at m/z 365. (**d**) Hydrolyzed and derivatized amino acids from Pcn1.7 (dashed line) and coinjected with DL-Lan standard (solid line). The trace was monitored at m/z 365. (**e**) Hydrolyzed and derivatized amino acids from Pcn1.7 (dashed line) and coinjected with LL-Lan standard (solid line). The trace was monitored at m/z 365. Lan and MeLan from Pcn1.7 used for overlays were adjusted to 70% intensity for clarity. A small amount of epimerization of the (Me)Lan occurs during acid hydrolysis of the peptides as has been reported previously (*21*).

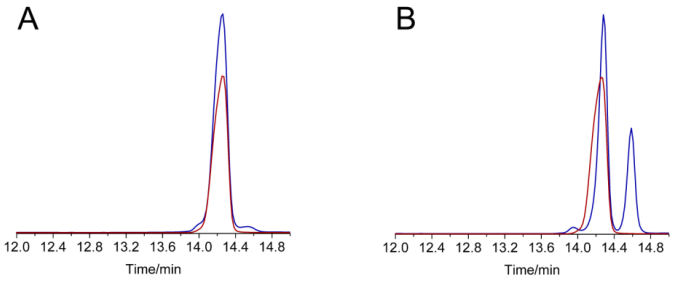

Figure 8.16 GC-MS traces (selected ion monitoring, SIM, at 379 Da for MeLan) of derivatized (Me)Lan obtained from Pcn1.1 and coinjections with derivatized Lan/MeLan standards. (**a**) Hydrolyzed and derivatized amino acids from Pcn1.1 (red line) and coinjected with DL-Melan standard with Pcn1.1 (blue line). (**b**) Hydrolyzed and derivatized amino acids from Pcn1.1 (red line) and coinjected with LL-Melan standard (blue line). Traces of derivatized MeLan from Pcn1.1 (red lines) used for overlay were adjusted to 70% intensity for clarity.

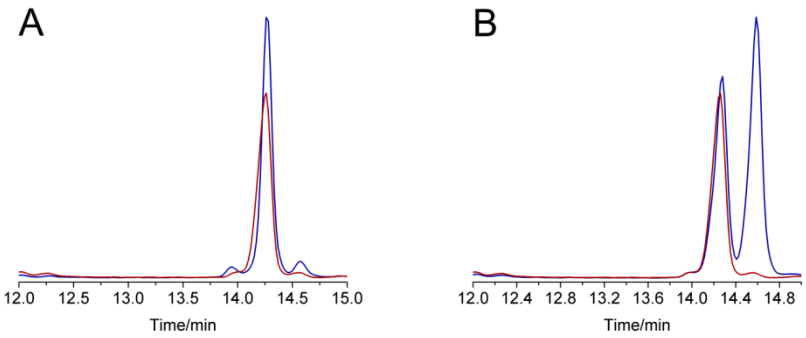

Figure 8.17 GC-MS traces (selected ion monitoring, SIM, at 379 Da for MeLan) of derivatized (Me)Lan obtained from Pcn3.3 and coinjections with derivatized Lan/MeLan standards.. (**a**) Hydrolyzed and derivatized amino acids from Pcn3.3 (red line) and coinjected with DL-MeLan standard (blue line). (**b**) Hydrolyzed and derivatized amino acids from Pcn3.3 (red line) and coinjected with LL-MeLan standard (blue line). Derivatized MeLan from Pcn3.3 used for overlays were adjusted to 70% intensity for clarity.

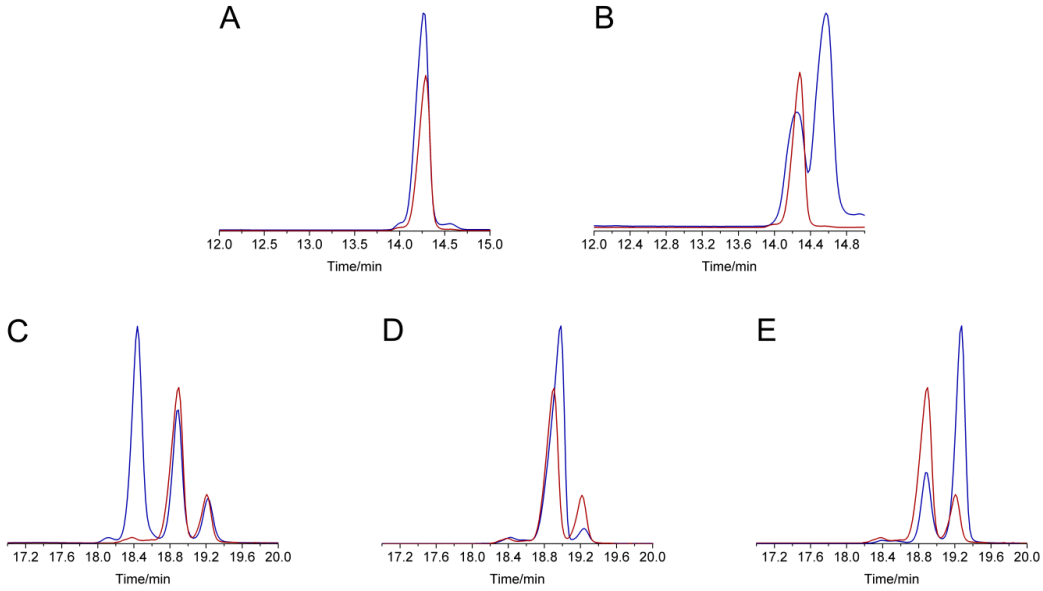

Figure 8.18 GC-MS traces (selected ion monitoring, SIM, 365 Da for Lan and 379 Da for MeLan) for coinjection of derivatized (Me)Lan standards and hydrolyzed and derivatized amino acids from Pcn4.3. (**a**) Hydrolyzed and derivatized amino acids from Pcn4.3 (red line) and coinjected with derivatized DL-MeLan standard (blue line). (**b**) Hydrolyzed and derivatized amino acids from Pcn4.3 (red line) and coinjected with derivatized LL-MeLan standard (blue line). (**c**) Hydrolyzed and derivatized amino acids from Pcn4.3 (red line) and coinjected with derivatized DD-Lan standard (blue line). (**d**) Hydrolyzed and derivatized amino acids from Pcn4.3 (red line) and coinjected with derivatized DL-Lan standard (blue line). (**e**) Hydrolyzed and derivatized amino acids from Pcn4.3 (red line) and coinjected with derivatized LL-Lan standard (blue line). Traces of derivatized Lan and MeLan from Pcn4.3 (red lines) used for overlay were adjusted to 70% intensity for clarity.

The GC trace for Pcn4.3, the other prochlorosin investigated here that contains Lan, was more complicated. Although the derivatized DL-Lan was still the dominant peak, an additional peak in the GC trace accounted for about 20% of the total Lan (**Figure 8.19a**). The material giving rise to this peak was confirmed to be derivatized LL-Lan by co-injections (**Figure 8.18**). This result was in agreement with the NMR data (1D water suppressed spectrum shown in **Figure 8.19b**), which exhibited minor peaks with integration values around 25% that of the main peaks. These NMR data demonstrate that the LL-Lan was already present in Pcn4.3 before acid hydrolysis and derivatization and was not introduced by epimerization during HCl hydrolysis.

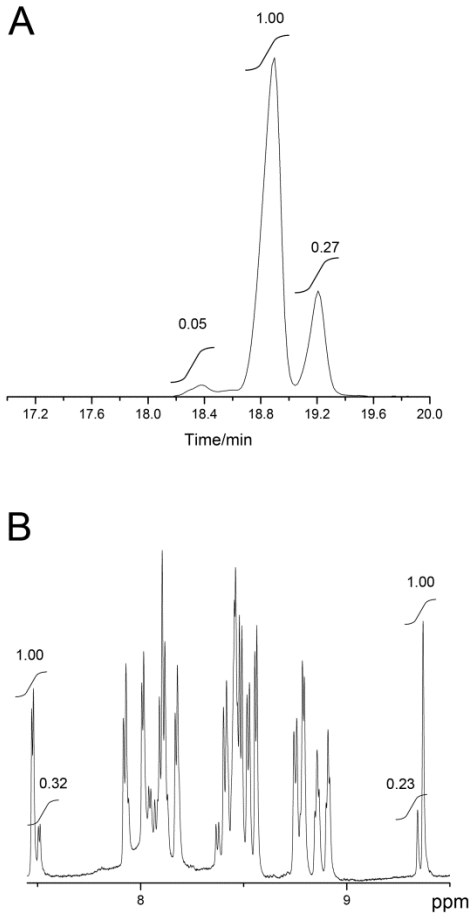

Figure 8.19 The LL-Lan diastereomer in Pcn4.3 and its relative abundance. (**a**) GC-MS trace of hydrolyzed and derivatized amino acids of Pcn4.3 monitored for Lan. DD-Lan (18.2-18.6 min), DL-Lan (18.6-19.0min) and LL-Lan (19.0-19.4 min) eluted from the column, with the relative peak areas indicated. (**b**) 1D water-suppressed NMR spectrum of Pcn4.3. Minor peaks originate from a different diastereomer of Pcn4.3 containing LL-Lan as shown in panel **a**. Integrations of two pairs of major/minor peaks are shown to estimate the relative amount of the diastereomer.

In addition to the prochlorosins for which the NMR structures were determined, the stereochemistry of the Lan and MeLan residues was determined for three other ProcA substrates that were posttranslationally modified by ProcM in *E. coli* (ProcA2.8, ProcA2.11, and ProcA3.2). The corresponding prochlorosins could not be produced in sufficient purity after leader peptide removal for NMR studies, but after hydrolysis and derivatization of the resulting amino acids, GC-MS analysis demonstrated that they contained Lan and MeLan residues with the same configurations as those in the other prochlorosins (**Figures 8.20-8.12**).

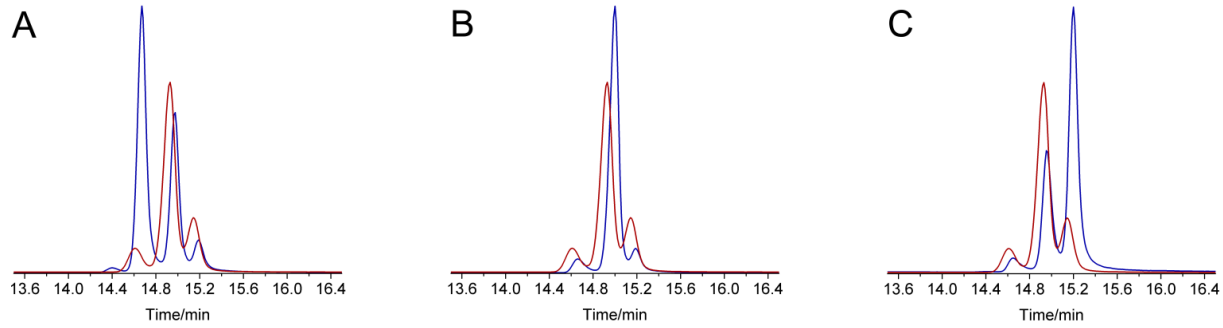

Figure 8.20 GC-MS traces (selected ion monitoring, SIM, 365 Da for Lan) for coinjection of derivatized (Me)Lan standards and hydrolyzed and derivatized amino acids from ProcM-modified ProcA2.8. (**a**) Hydrolyzed and derivatized amino acids from ProcM-modified ProcA2.8 (red line) and coinjected with derivatized DD-Lan standard (blue line). (**b**) Hydrolyzed and derivatized amino acids from ProcM-modified ProcA2.8 (red line) and coinjected with derivatized DL-Lan standard (blue line). (**c**) Hydrolyzed and derivatized amino acids from ProcM-modified ProcA2.8 (red line) and coinjected with derivatized LL-Lan standard (blue line). Traces of derivatized Lan and MeLan from ProcM-modified ProcA2.8 (red lines) used for overlay were adjusted to 70% intensity for clarity. A small amount of epimerization of the (Me)Lan occurs during acid hydrolysis of the peptides as has been reported previously (*21*).

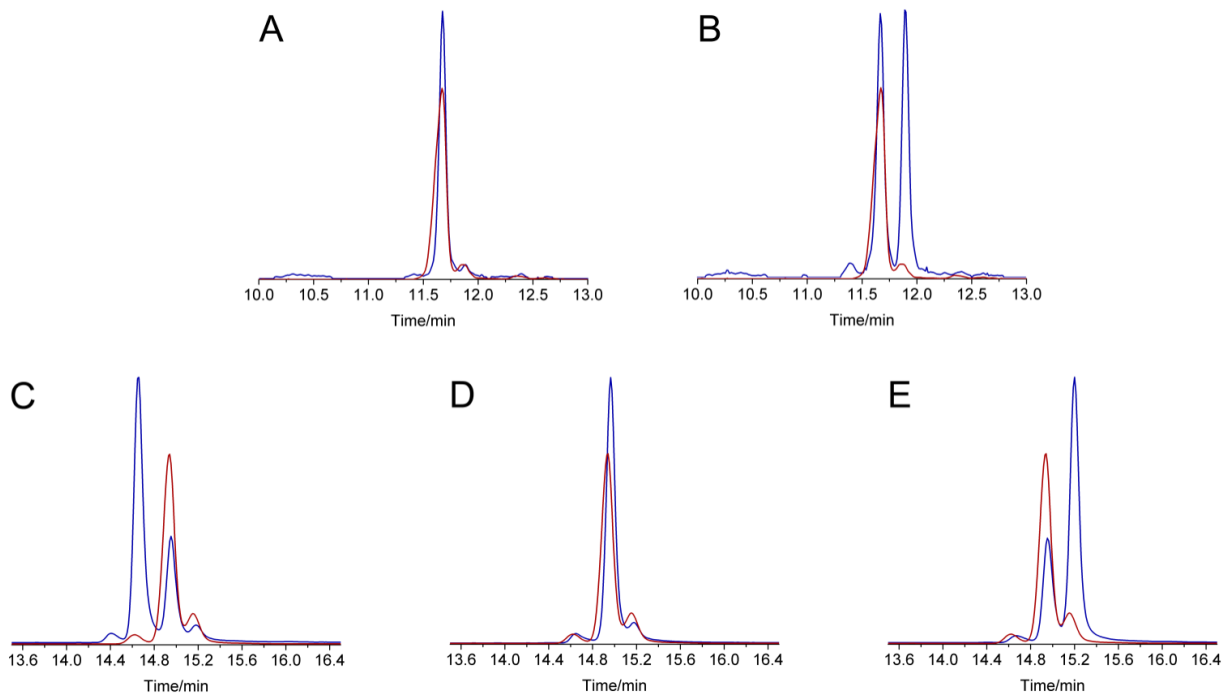

Figure 8.21 GC-MS traces (selected ion monitoring, SIM, 365 Da for Lan and 379 Da for MeLan) for coinjection of derivatized (Me)Lan standards and hydrolyzed and derivatized amino acids from ProcM-modified ProcA2.11. (**a**) Hydrolyzed and derivatized amino acids from ProcM-modified ProcA2.11 (red line) and coinjected with derivatized DL-MeLan standard (blue line). (**b**) Hydrolyzed and derivatized amino acids from ProcM-modified ProcA2.11 (red line) and coinjected with derivatized LL-MeLan standard (blue line). (**c**) Hydrolyzed and derivatized amino acids from ProcM-modified ProcA2.11 (red line) and coinjected with derivatized DD-Lan standard (blue line). (**d**) Hydrolyzed and derivatized amino acids from ProcM-modified ProcA2.11 (red line) and coinjected with derivatized DL-Lan standard (blue line). (**e**) Hydrolyzed and derivatized amino acids from ProcM-modified ProcA2.11 (red line) and coinjected with derivatized LL-Lan standard (blue line). Traces of derivatized Lan and MeLan from ProcM-modified ProcA2.11 (red lines) used for overlay were adjusted to 70% intensity for clarity. A small amount of epimerization of the (Me)Lan occurs during acid hydrolysis of the peptides as has been reported previously (*21*).

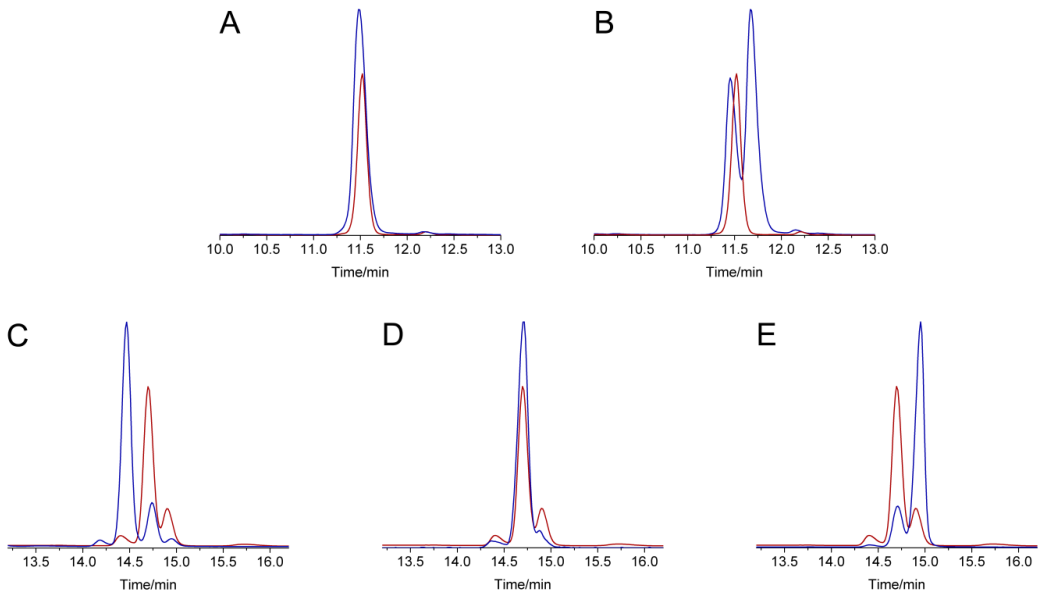

Figure 8.22 GC-MS traces (selected ion monitoring, SIM, 365 Da for Lan and 379 Da for MeLan) for coinjection of derivatized (Me)Lan standards and hydrolyzed and derivatized amino acids from ProcM-modified ProcA3.2. (**a**) Hydrolyzed and derivatized amino acids from ProcM-modified ProcA3.2 (red line) and coinjected with derivatized DL-MeLan standard (blue line). (**b**) Hydrolyzed and derivatized amino acids from ProcM-modified ProcA3.2 (red line) and coinjected with derivatized LL-MeLan standard (blue line). (**c**) Hydrolyzed and derivatized amino acids from ProcM-modified ProcA3.2 (red line) and coinjected with derivatized DD-Lan standard (blue line). (**d**) Hydrolyzed and derivatized amino acids from ProcM-modified ProcA3.2 (red line) and coinjected with derivatized DL-Lan standard (blue line). (**e**) Hydrolyzed and derivatized amino acids from ProcM-modified ProcA3.2 (red line) and coinjected with derivatized LL-Lan standard (blue line). Traces of derivatized Lan and MeLan from ProcM-modified ProcA3.2 (red lines) used for overlay were adjusted to 70% intensity for clarity. A small amount of epimerization of the (Me)Lan occurs during acid hydrolysis of the peptides as has been reported previously (*21*).

8.3 Discussion

In this study I set out to investigate three aspects of prochlorosins biosynthesis: whether tandem MS is reliable for determination of ring topologies, whether sufficient quantities of material can be produced such that NMR spectroscopy can be used when MS data are inconclusive, and whether the Lan and MeLan residues in prochlorosins adopt the canonical stereochemistry as was found in most other lanthipeptides.

To address these questions, I first produced large amounts of prochlorosins heterologously because less than 10 μg of impure products were obtained from 20 L of culture of

Prochlorococcus MIT9313 (*1*). Although the peptides after proteolysis are not completely identical to the mature prochlorosins, for determination of the ring topology and the stereochemistry of (Me)Lan residues, the additional five amino acids of the leader peptide are inconsequential.

The NMR data reported here confirm the ring topologies of three prochlorosins that were previously proposed on the basis of tandem mass spectrometry in combination with site-directed mutagenesis. Thus, the concern that a highly promiscuous enzyme like ProcM might process a mutant peptide differently from a wild-type peptide appears not to be corroborated for these compounds. This concern arose when the B ring of Pcn4.3 could not be determined by tandem MS in a previous study (*1*). ProcA4.3 is dehydrated three times and the previously observed fragmentation pattern indicated that Thr1 and Ser3 were converted to Dhb1 and Dha3, respectively. Whether the third dehydrated residue was Thr8 or Ser9 could not be determined. Similarly, although the tandem MS data demonstrated that Pcn4.3 contained two non-overlapping rings, whether Thr8 or Ser9 was converted to a thioether crosslink in the B ring could not be deduced (*1*). Typically, such questions have been addressed by mutating the Thr and Ser in question (*1, 22, 23*). If Thr8 is dehydrated in wt ProcA4.3, mutation of Thr8 to Ala was expected to result in two instead of three dehydrations. Conversely, if Ser9 was dehydrated in wt ProcA4.3, mutation of Thr8 to Ala should not result in a difference in the number of dehydrations. This latter result was observed as the ProcA4.3A-Thr8Ala mutant was dehydrated three times and contained two Lan rings (*1*). However, when Ser9 was then mutated to Ala to confirm this result, two dehydrations were expected but three were still observed. In other words, at least one of the mutants must be processed differently by ProcM than wt ProcA4.3. Therefore, the question whether Thr8 or Ser9 is involved in the B ring could not be answered. The NMR data in this study show conclusively that it is Thr8 that is dehydrated in wt ProcA4.3 and that it makes a thioether crosslink with Cys12 (**Figure 8.14**).

Collectively, these results raise the question why mutants of ProcA1.7 and ProcA3.3 that were made in the previous study to deduce their ring topologies by tandem MS were processed correctly by ProcM, as shown by the NMR data in this study, whereas mutants of ProcA4.3 were not. Although this question cannot be addressed directly, processing of ProcA4.3 by ProcM appears to be less robust in general. Firstly, mutations at the −1 position in the precursor peptide

completely abolished ProcM activity in the current study whereas ProcM fully processed mutants at the −1 position for the other three ProcA substrates. Secondly, whereas all Lan and MeLan residues in Pcn1.1, Pcn1.7, and Pcn3.3 were formed in high stereochemical purity with the same stereochemistry (DL), the Lan in Pcn4.3 was decidedly less pure and contained about 20% of the LL diastereomer. As shown in **Figure 8.14**, the A ring of Pcn4.3 contains two Gly residues between the Dhb and Cys that form this lanthionine and perhaps the resulting increased conformational freedom allows enzymatic cyclization that results in protonation of the enolate from the opposite face. Alternatively, the lower stereochemical fidelity of cyclization of this particular Lan structure could be the result of non-enzymatic cyclization as discussed below.

The determination of the ring topologies of four different prochlorosins in this study demonstrates the remarkable substrate flexibility of the ProcM enzyme. Inspection of **Figure 8.14** illustrates that the structures of the four selected prochlorosins do not exhibit any similarity to each other nor to any other known lantibiotics/lanthipeptides (*4*). The rings generated by ProcM contain very different numbers of amino acids, from four residues in the A rings of Pcn1.7 and Pcn4.3 to 11 residues in the A ring of Pcn3.3. Furthermore, the rings are generated either from a Cys that is located N-terminally to the dehydro amino acid with which it reacts (A ring of Pcn 1.7 and B ring of Pcn3.3) or from a Cys located C-terminally of its partner dehydro amino acid (all other rings). Finally, Pcn1.1 and 4.3 have non-overlapping rings, Pcn1.7 has overlapping rings, and Pcn3.3 contains a ring within a ring. Although many lantipeptide cyclases achieve formation of different sized rings and with different topologies, ProcM, and presumably its many orthologs in the world's oceans (*1*), is unique in that it acts on so many different substrates.

It was this unprecedented flexibility that led to a proposal that perhaps the enzyme generates only a subset of the rings and that the partially cyclized intermediates would have a preorganized structure for non-enzymatic cyclization of the remaining rings (*1*). This proposal was bolstered by the previously observed facile non-enzymatic intramolecular addition of Cys to dehydro amino acids located N-terminal to the Cys residue (*24-30*). These model studies on non-enzymatic cyclization showed that formation of individual small rings is often stereoselective and results in the same stereochemistry as that observed for enzymatic cyclization. However, for larger rings or multiple ring structures, non-enzymatic cyclization has been shown to give

mixtures of diasteromers and/or constitutional isomers (*26, 29, 31*), and cyclization involving a Cys located N-terminally to its partner dehydro amino acid has been shown to be non-stereoselective (*28*). Therefore, in the current investigation the stereochemistry of the Lan and MeLan structures of prochlorosins was determined to test for non-enzymatic cyclization. The GC-MS data clearly illustrate that the Lan residues in prochlorosins have the (2*S*, 6*R*) configuration and that the MeLan residues have the (2*S*, 3*S*, 6*R*) configuration. These configurations are the same as those found for most lantibiotics for which the stereochemistry has been determined (*7*). As discussed above, the only exception was the Lan in Pcn4.3, which was made up of a 4:1 ratio of DL and LL diastereomers. Although these results do not rule out non-enzymatic cyclization, they also do not provide support for non-enzymatic cyclization and based on the data in this study on seven prochlorosins with very different ring structures, a model in which ProcM does generate all thioether crosslinks is favored. This model is further supported by a recent report in which the non-enzymatic cyclization of a few dehydrated ProcA peptides was confirmed to be much slower than ProcM-catalyzed ring cyclization (*32*). One interesting unique aspect of ProcM is that based on sequence alignments, its catalytic Zn^{2+} is liganded by three Cys residues rather than the two Cys and one His residue that is found for all other LanM and LanC lanthionine cyclases investigated thus far (*33-37*). The Zn^{2+} has been proposed to activate the Cys residues of the substrate for nucleophilic attack onto the dehydro amino acids (*33*). Model studies on activation of thiolate nucleophiles by Zn^{2+} have demonstrated increased reactivity with increased number of thiolate ligands (*38-41*). Therefore, ProcM may derive its promiscuity in part from a highly activating Zn^{2+} site with three Cys ligands from ProcM and one Cys ligand from the substrate (*42*).

In summary, the structures of four prochlorosins were determined by NMR spectroscopy in this chapter. I further show for seven prochlorosins that they contain Lan and MeLan residues of the same configuration as those found in other lantibiotics/lanthipeptides. These data suggest that most, if not all, of the rings are generated by the ProcM enzyme despite the remarkable structural diversity of its products.

8.4 Methods

General methods

All polymerase chain reactions (PCR) were carried out on a C1000 thermal cycler (Bio-Rad). DNA sequencing was performed by ACGT, Inc. Preparative HPLC was performed using a Waters Delta 600 instrument equipped with appropriate columns. Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) was carried out on a Voyager-DE-STR (Applied Biosystems). Overexpression and purification of the modified precursor peptides were performed using previously reported procedures (*12*).

Materials

All oligonucleotides were purchased from Integrated DNA Technologies and used as received. Restriction endonucleases, DNA polymerases and T4 DNA ligase were from New England Biolabs. Media components were obtained from Difco laboratories. *E. coli* DH5α and *E. coli* BL21 (DE3) cells were used as host for cloning and plasmid propagation and host for expression, respectively. Endoproteinases were ordered from Roche Biosciences. Chemicals were purchased from Sigma Aldrich, Fisher Scientific or CalBiochem.

Construction of pRSFDuet-1 derivatives for coexpression of ProcM and ProcAs

The genes encoding ProcA1.1G−1E and ProcA4.3 were cloned from previously reported pET15b plasmids containing these genes (*1*) and inserted into the multiple cloning site 1 of pRSFDuet-1/procM-2 plasmid (*12*) using the *EcoR*I and *Not*I restriction sites. Primer sequences used are shown in **Table 8.1**. Negative numbers are used for amino acids in the leader peptide counting backwards from the putative leader peptide cleavage site. The sequences of the constructed plasmids were confirmed by DNA sequencing. The pRSFDuet-1/procM-2, pRSFDuet-1/procA3.3G−1K/procM-2, pRSFDuet-1/procA1.7G−1R/procM-2, pRSFDuet-1/procA2.8/procM-2, pRSFDuet-1/procA2.11G−1K/procM-2, pRSFDuet-1/procA3.2/procM-2, pRSFDuet-1/procA4.3G−1R/procM-2 and pRSFDuet-1/procA4.3G−1E/procM-2 plasmids were obtained as reported previously (*12*).

Table 8.1 Primer sequences for cloning of ProcAs.

Primer Name	Primer Sequence (5'-3')
ProcA1.1_EcoRI_FP_Duet	AAG AAT TCG ATG AAA AAG CGA CTC AAC
ProcA1.1_NotI_RP_Duet	AAG CGG CCG CTC AGC ACA CAT TGA TAG
ProcA4.3_EcoRI_FP_Duet	GGT GAG TGG AAT TCG ATG TCA GAA GAA CAA CTG AAG GC
ProcA4.3_NotI_RP_Duet	ATG ACC TAG CGG CCG CCT AGT AAC AAA ACA TAC

Protease cleavage and purification of prochlorosin core peptides

Post-translationally modified ProcA peptides were treated with the commercial protease GluC to remove the leader peptides. For some peptides, other proteases were used to hydrolyze the leader peptides into smaller fragments which sometimes proved advantageous for purification (LysC for ProcA1.1G−1E and ProcA3.3G−1K, trypsin for ProcA4.3). In such instances, the modified core peptide was not proteolyzed even if it contained potential cleavage sites for LysC/trypsin, presumably because the presence of dehydrated amino acids and/or (Me)Lan residues greatly reduced the proteolytic efficiency. The protease cleavage reactions were quenched with 0.5% TFA, and the desired products were confirmed by MALDI-TOF MS and purified by reversed phase (RP) HPLC using a Jupiter proteo C12 column (5 μm; 90 Å; 250 x 10.0 mm). Solvents for RP-HPLC were solvent A (0.1% TFA in water) and solvent B (0.086% TFA in 80% acetonitrile/20% water). The desired core peptides eluted from the column between 40 and 60% solvent B and pure fractions were lyophilized.

NMR spectroscopy

NMR spectra were acquired at the NMR Facility (School of Chemical Sciences, University of Illinois at Urbana-Champaign) on either a Varian INOVA 600 MHz or 500 MHz spectrometer equipped with a 5 mm triple resonance (^{1}H-^{13}C-^{15}N) triaxial gradient probe. Lyophilized peptides were dissolved in 90% H_2O/10% D_2O to a final concentration between 5 and 7 mM. NMR spectra were acquired at 20 ^{o}C for 2048 direct and 256/400 indirect data points with 16 or 32 scans depending on the peptide concentrations. TOCSY (*43*) (mixing time 0.080 s), gradient double quantum filtered COSY (gDQCOSY) (*44*) and NOESY (*45*) (mixing time 0.20 s) spectra were acquired with solvent suppression by presaturation or water gate pulse sequences. The lyophilized peptides were dissolved in 100% D_2O and NOESY spectra were acquired without

exchangeable proton signals. NOESY spectra of Pcn 1.1 and 3.3 were acquired with a mixing time of 0.20 s, while those of Pcn1.7 and 4.3 were acquired with 0.40 s for clearer NOE signals. Spectra were processed with NMRPipe (*46*) and analyzed in Sparky (*47*).

GC-MS analysis

The synthesis of Lan and MeLan standards and the preparation of samples for GC-MS analysis were carried out following a reported procedure (*10*). For Pcn1.1, 1.7, 3.3, and 4.3, the purified modified core peptides were used. For Pcn2.8, 2.11, and 3.2, the ProcM-modified precursor peptides with their leader sequences still attached were used for amino acid analysis. The derivatized samples were analyzed by GC-MS using an Agilent HP 6890N mass spectrometer equipped with a Varian CP-Chirasil-L-Val fused silica column (25 m x 0.25 mm x 0.15 μm). Samples were dissolved in methanol and introduced to the instrument via a splitless injection at a flow rate of 1.7 or 2.0 mL/min helium gas. The temperature method used was 160 ^{o}C for 5 min, 160 ^{o}C to 180 ^{o}C at 3 ^{o}C/min, and 180 ^{o}C for 10 min. The mass spectrometer was operated in simultaneous scan/selected ion monitoring (SIM) mode, monitoring at the characteristic fragment masses of 365 Da for Lan and 379 Da for MeLan residues.

8.5 References

1. Li, B., Sher, D., Kelly, L., Shi, Y., Huang, K., Knerr, P. J., Joewono, I., Rusch, D., Chisholm, S. W., and van der Donk, W. A. (2010) Catalytic promiscuity in the biosynthesis of cyclic peptide secondary metabolites in planktonic marine cyanobacteria, *Proc. Natl. Acad. Sci. U.S.A. 107*, 10430-10435.
2. Zhang, Q., Yang, X., Wang, H., and van der Donk, W. A. (2014) High divergence of the precursor peptides in combinatorial lanthipeptide biosynthesis, *ACS Chem. Biol. 9*, 2686-2694.
3. Johnson, Z. I., Zinser, E. R., Coe, A., McNulty, N. P., Woodward, E. M., and Chisholm, S. W. (2006) Niche partitioning among Prochlorococcus ecotypes along ocean-scale environmental gradients, *Science 311*, 1737-1740.
4. Willey, J. M., and van der Donk, W. A. (2007) Lantibiotics: peptides of diverse structure and function, *Annu. Rev. Microbiol. 61*, 477-501.
5. Haft, D. H., Basu, M. K., and Mitchell, D. A. (2010) Expansion of ribosomally produced natural products: a nitrile hydratase- and Nif11-related precursor family, *BMC Biol. 8*, 70.
6. Xie, L., Miller, L. M., Chatterjee, C., Averin, O., Kelleher, N. L., and van der Donk, W. A. (2004) Lacticin 481: in vitro reconstitution of lantibiotic synthetase activity, *Science 303*, 679-681.
7. Chatterjee, C., Paul, M., Xie, L., and van der Donk, W. A. (2005) Biosynthesis and mode of action of lantibiotics, *Chem. Rev. 105*, 633-684.
8. Tang, W., and van der Donk, W. A. (2013) The sequence of the enterococcal cytolysin imparts unusual lanthionine stereochemistry, *Nat. Chem. Biol. 9*, 157-159.
9. Küsters, E., Allgaier, H., Jung, G., and Bayer, E. (1984) Resolution of sulphur-containing amino acids by chiral phase gas chromatography, *Chromatographia 18*, 287-293.
10. Liu, W., Chan, A. S., Liu, H., Cochrane, S. A., and Vederas, J. C. (2011) Solid supported chemical syntheses of both components of the lantibiotic lacticin 3147, *J. Am. Chem. Soc. 133*, 14216-14219.
11. Knerr, P. J., Tzekou, A., Ricklin, D., Qu, H., Chen, H., van der Donk, W. A., and Lambris, J. D. (2011) Synthesis and activity of thioether-containing analogues of the complement inhibitor compstatin, *ACS Chem. Biol. 6*, 753-760.
12. Shi, Y., Yang, X., Garg, N., and van der Donk, W. A. (2011) Production of lantipeptides in *Escherichia coli*, *J. Am. Chem. Soc. 133*, 2338-2341.
13. Ökesli, A., Cooper, L. E., Fogle, E. J., and van der Donk, W. A. (2011) Nine post-translational modifications during the biosynthesis of cinnamycin, *J. Am. Chem. Soc. 133*, 13753-13760.

14. Velásquez, J. E., Zhang, X., and van der Donk, W. A. (2011) Biosynthesis of the Antimicrobial Peptide Epilancin 15X and its Unusual N-terminal Lactate Moiety, *Chem. Biol. 18*, 857-867.
15. Freund, S., Jung, G., Gibbons, W. A., and Sahl, H. G. (1991) NMR and circular dichroism studies on Pep5, *Nisin and Novel Lantibiotics*, 103-112.
16. Van de Ven, F. J., Van den Hooven, H. W., Konings, R. N., and Hilbers, C. W. (1991) NMR studies of lantibiotics. The structure of nisin in aqueous solution, *Eur. J. Biochem. 202*, 1181-1188.
17. Martin, N. I., Sprules, T., Carpenter, M. R., Cotter, P. D., Hill, C., Ross, R. P., and Vederas, J. C. (2004) Structural characterization of lacticin 3147, a two-peptide lantibiotic with synergistic activity, *Biochemistry 43*, 3049-3056.
18. Shenkarev, Z. O., Finkina, E. I., Nurmukhamedova, E. K., Balandin, S. V., Mineev, K. S., Nadezhdin, K. D., Yakimenko, Z. A., Tagaev, A. A., Temirov, Y. V., Arseniev, A. S., and Ovchinnikova, T. V. (2010) Isolation, structure elucidation, and synergistic antibacterial activity of a novel two-component lantibiotic lichenicidin from *Bacillus licheniformis* VK21, *Biochemistry 49*, 6462-6472.
19. Smith, L., Novak, J., Rocca, J., McClung, S., Hillman, J. D., and Edison, A. S. (2000) Covalent structure of mutacin 1140 and a novel method for the rapid identification of lantibiotics, *Eur. J. Biochem. 267*, 6810-6816.
20. He, Z., Yuan, C., Zhang, L., and Yousef, A. E. (2008) N-terminal acetylation in paenibacillin, a novel lantibiotic, *FEBS Lett. 582*, 2787-2792.
21. Kido, Y., Hamakado, T., Yoshida, T., Anno, M., Motoki, Y., Wakamiya, T., and Shiba, T. (1983) Isolation and characterization of ancovenin, a new inhibitor of angiotensin I converting enzyme, produced by actinomycetes, *J. Antibiot. 36*, 1295-1299.
22. Cooper, L. E., McClerren, A. L., Chary, A., and van der Donk, W. A. (2008) Structure-activity relationship studies of the two-component lantibiotic haloduracin, *Chem. Biol. 15*, 1035-1045.
23. Caetano, T., Krawczyk, J. M., Mosker, E., Süssmuth, R. D., and Mendo, S. (2011) Heterologous expression, biosynthesis, and mutagenesis of type II lantibiotics from *Bacillus licheniformis* in *Escherichia coli*, *Chem. Biol. 18*, 90-100.
24. Polinsky, A., Cooney, M. G., Toy-Palmer, A., Ösapay, G., and Goodman, M. (1992) Synthesis and conformational properties of the lanthionine-bridged opioid peptide [D-Ala_L^2,Ala_L^2]enkephalin as determined by NMR and computer simulations, *J. Med. Chem. 35*, 4185-4191.
25. Toogood, P. L. (1993) Model Studies of Lantibiotic Biogenesis, *Tetrahedron Lett. 34*, 7833-7836.

26. Burrage, S., Raynham, T., Williams, G., Essex, J. W., Allen, C., Cardno, M., Swali, V., and Bradley, M. (2000) Biomimetic synthesis of lantibiotics, *Chem.-Eur. J. 6*, 1455-1466.
27. Okeley, N. M., Zhu, Y., and van der Donk, W. A. (2000) Facile chemoselective synthesis of dehydroalanine-containing peptides, *Org. Lett. 2*, 3603-3606.
28. Zhou, H., and van der Donk, W. A. (2002) Biomimetic stereoselective formation of methyllanthionine, *Org. Lett. 4*, 1335-1338.
29. Zhu, Y., Gieselman, M., Zhou, H., Averin, O., and van der Donk, W. A. (2003) Biomimetic studies on the mechanism of stereoselective lanthionine formation, *Org. Biomol. Chem. 1*, 3304-3315.
30. Zhang, X., Ni, W., and van der Donk, W. A. (2007) On the regioselectivity of thioether formation by lacticin 481 synthetase, *Org. Lett. 9*, 3343-3346.
31. Zhang, X., and van der Donk, W. A. (2007) On the substrate specificity of the dehydratase activity of lacticin 481 synthetase, *J. Am. Chem. Soc. 129*, 2212 - 2213.
32. Mukherjee, S., and van der Donk, W. A. (2014) Mechanistic Studies on the Substrate-Tolerant Lanthipeptide Synthetase ProcM, *J. Am. Chem. Soc. 136*, 10450-10459.
33. Okeley, N. M., Paul, M., Stasser, J. P., Blackburn, N., and van der Donk, W. A. (2003) SpaC and NisC, the cyclases involved in subtilin and nisin biosynthesis, are zinc proteins, *Biochemistry 42*, 13613-13624.
34. Li, B., Yu, J. P., Brunzelle, J. S., Moll, G. N., van der Donk, W. A., and Nair, S. K. (2006) Structure and mechanism of the lantibiotic cyclase involved in nisin biosynthesis, *Science 311*, 1464-1467.
35. Helfrich, M., Entian, K. D., and Stein, T. (2007) Structure-function relationships of the lanthionine cyclase SpaC involved in biosynthesis of the *Bacillus subtilis* peptide antibiotic subtilin, *Biochemistry 46*, 3224-3233.
36. Li, B., and van der Donk, W. A. (2007) Identification of essential catalytic residues of the cyclase NisC involved in the biosynthesis of nisin, *J. Biol. Chem. 282*, 21169-21175.
37. Paul, M., Patton, G. C., and van der Donk, W. A. (2007) Mutants of the zinc ligands of lacticin 481 synthetase retain dehydration activity but have impaired cyclization activity, *Biochemistry 46*, 6268-6276.
38. Wilker, J. J., and Lippard, S. J. (1997) Alkyl transfer to metal thiolates: kinetics, active species identification, and relevance to the DNA methyl phosphotriester repair center of *Escherichia coli* Ada, *Inorg. Chem. 36*, 969-978.
39. Warthen, C. R., Hammes, B. S., Carrano, C. J., and Crans, D. C. (2001) Methylation of neutral pseudotetrahedral zinc thiolate complexes: model reactions for alkyl group transfer to sulfur by zinc-containing enzymes, *J. Biol. Inorg. Chem. 6*, 82-90.

40. Brand, U., Rombach, M., Seebacher, J., and Vahrenkamp, H. (2001) Functional modeling of cobalamine-independent methionine synthase with pyrazolylborate-zinc-thiolate complexes, *Inorg. Chem. 40*, 6151-6157.
41. Chiou, S. J., Riordan, C. G., and Rheingold, A. L. (2003) Synthetic modeling of zinc thiolates: Quantitative assessment of hydrogen bonding in modulating sulfur alkylation rates, *Proc. Natl. Acad. Sci. U. S. A. 100*, 3695-3700.
42. Yu, Y., Mukherjee, S., and van der Donk, W. A. (2015) Product formation by the promiscuous lanthipeptide synthetase ProcM is under kinetic control, *J. Am. Chem. Soc. accepted*.
43. Braunschweiler, L., and Ernst, R. R. (1983) Coherence transfer by isotropic mixing - application to proton correlation spectroscopy, *J. Magn. Reson. 53*, 521-528.
44. Piantini, U., Sorensen, O. W., and Ernst, R. R. (1982) Multiple quantum filters for elucidating NMR coupling networks, *J. Am. Chem. Soc. 104*, 6800-6801.
45. Jeener, J., Meier, B. H., Bachmann, P., and Ernst, R. R. (1979) Investigation of exchange processes by 2-dimensional NMR spectroscopy, *J. Chem. Phys. 71*, 4546-4553.
46. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes, *J. Biomol. NMR 6*, 277-293.
47. Goddard, T. D., and Kneeler, D. G. (2005) Sparky, University of California, San Francisco.